Institute of Education

The role of background knowledge in reading comprehension of subject-specific texts

Alenka Umek

Thesis submitted for the degree of Doctor of Philosophy

Institute of Education
University College London

2023

DECLARATION

I, Alenka Umek, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Word count (exclusive of references and appendices): 81,979 words

ABSTRACT

This thesis investigates the impact of background knowledge in L2 reading comprehension of subject-specific texts, in particular its interaction with grammar knowledge. It explores how different levels of discipline-related background knowledge, grammar knowledge, and self-reported familiarity affect individual differences in L2 reading comprehension in terms of its outcomes and process. A number of studies made assumptions about readers' knowledge based on their study disciplines or reports by readers themselves, and thus this study also explores the difference between two operationalizations: tested background knowledge and self-reported familiarity.

A mixed-methods approach was used by combining two studies: a testing study and a think-aloud study. Altogether 404 students of the School of Economics and Business, University of Ljubljana, Slovenia took part in the study; 22 in the piloting study and 382 in the main study, out of which 358 were engaged in the testing study and 24 in the think-aloud study. The quantitative and qualitative datasets were obtained from five research instruments: a grammar test, a test of discipline-related background knowledge, a reading comprehension test based on three finance texts, a post-reading questionnaire, and think-aloud verbal protocols.

The results of multiple regression revealed that tested background knowledge was a significant medium strength predictor of reading comprehension, slightly stronger than grammar knowledge. In contrast, self-reported familiarity was not found to impact reading comprehension and was not its predictor. This evidence casts doubt over self-reporting as an operationalization of knowledge in L2 reading.

Apart from having a facilitative effect on L2 reading comprehension, background knowledge was found to have compensatory and additive roles when interacting with grammar knowledge. The findings showed that readers with higher discipline-related background knowledge could use it to make up for lower grammar knowledge and vice versa, thus suggesting the compensation effect between the two variables. In addition, the results revealed that readers were able to use their background knowledge regardless of their level of grammar knowledge, albeit slightly less at higher levels of grammar knowledge. This finding suggests that the threshold hypothesis could not be supported. Finally, the group of students with both high

background knowledge and high grammar knowledge outperformed other groups in reading comprehension, which suggests that the two variables affect reading comprehension in an additive way.

The qualitative data from verbal protocols in the think-aloud study and readers' scores from the testing study were obtained to compare the processing patterns and strategies used by readers with high and low background knowledge. Although both groups were found to use the same types of strategies, they differed in the frequency of their use. While the high background knowledge group used more correct paraphrases, elaboration, inferences, and evaluating, the low knowledge group adopted a more local-level approach by paying more attention to individual words, phrases, and sentences and reporting on various comprehension problems and inability to see the bigger picture. The results suggest differences between the groups with different levels of background knowledge with regard to semantic and pragmatic processing at the local and global level. Analysis of the verbal protocol and post-questionnaire data revealed that specialist vocabulary was the main source of difficulty in L2 reading comprehension of subject-specific texts.

IMPACT STATEMENT

Given the central importance of L2 reading for academic and professional success in various disciplines, this thesis explores the role of discipline-related background knowledge in L2 reading comprehension of subject-specific texts. It is underpinned by theoretical and empirical research in applied linguistics, especially ESL teaching, as well as cognitive psychology.

The thesis contributes to research by providing a better insight into the role of discipline-related background knowledge and grammar knowledge in L2 reading. It specifies the relative contributions of these variables and their predictive value in L2 reading comprehension. It empirically confirms the significant facilitative role of discipline-related background knowledge and suggests that having text-relevant background knowledge of the discipline benefits comprehension and works in tandem with grammar knowledge. The findings also highlight the compensatory and complementary relationship between discipline-related background knowledge and grammar knowledge in L2 reading comprehension.

The study provides empirical evidence for the distinction between tested background knowledge and self-reported familiarity with regard to their effect on L2 reading comprehension of subject-specific texts. While tested background knowledge had a statistically significant impact of medium strength on L2 reading comprehension, self-reported familiarity did not have any significant impact. This finding adds empirical grounds for a critical consideration of using self-reported topic familiarity as knowledge operationalization in future research. It suggests rethinking the operationalization and measurement of background knowledge in L2 reading research, especially when it involves discipline-related knowledge.

This study contributes to the scarce mixed-method research into the role of discipline-related knowledge in L2 reading. Among the implications for instruction, it highlights the central role of specialist technical vocabulary, which turned out to be a major reading difficulty in verbal protocols and students' comments in post-reading questionnaires. As learning terminology is at the heart of learning a discipline, the study suggests that instruction ought to provide ample opportunities for the readers to see how discipline-related knowledge is communicated through the vocabulary, structure, and structural patterns of expository texts and thus develop readers' disciplinary literacy.

ACKNOWLEDGEMENTS

This thesis would not have been written without the support and assistance of a number of people, to whom I am very grateful.

I would like first to thank my supervisor Amos Paran for his guidance and patient support at all stages of my research. I appreciated his comments and intellectual generosity.

I am grateful to students at the School of Economics and Business University of Ljubljana who took part in the preliminary and main studies of this research, especially those students who took time to participate in the think-aloud study. My thanks go to my colleagues at the Academic Unit for Languages for their input at the validation of grammar test. Many thanks are due to my colleagues at the Academic Unit for Money and Finance who were involved in the text selection and item writing of the reading comprehension test and background knowledge test. A very special thank you goes to Aljoša Valentinčič, whose input in initial text selection and validation of instruments was greatly appreciated.

My greatest thanks go to Sebastian Kočar and Mojca Bavdaž Lindič, who have most generously shared their knowledge of statistics; without their help this project would not have been completed. A sincere thank you goes to Dawn Reindl for proofreading.

I remember with gratitude my teachers and colleagues from the University of Reading with whom I discussed this project at its very beginning: the late Cyril J. Weir, Barry O'Sullivan, and Rita Green.

Thanks are due to my employer, the School of Economics and Business University of Ljubljana, for support and partial funding, which made my burden slightly lighter.

Finally, my heartfelt thanks go to my family and friends for their steadfast belief that I would finish this almost-abandoned project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
BK	Background knowledge
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ESL	English as a Second Language
ESP	English for Specific Purposes
GK	Grammar knowledge
L1	First language
L2	Second language
LP	Language proficiency
MCQ	Multiple choice questions
PRQ	Post-reading questionnaire
RC	Reading comprehension
SAQ	Short answer questions
SRF	Self-reported familiarity
TA	Think aloud
TEAP	Test of English for Academic Purposes
TEEP	Test of English for Educational Purposes
TOEFL	Test of English as a Foreign Language

CHAPTER 1
INTRODUCTION

1.1 Introduction

Reading in a second language (L2) is a skill of growing importance for academic and professional development. L2 reading research has examined various contributors to effective reading within disciplines. It has established that L2 proficiency and background knowledge are important factors in effective reading (Clapham, 1996; Chen & Donin, 1998; Uso-Juan, 2006; Al Shumaimeri, 2006; Lin & Chern, 2014; Karimi, 2017). However, there are conflicting views regarding the relative contributions of both factors and the ways their effect can be observed in the L2 reading process. There is a lack of research looking into the various knowledge operationalizations and how they compare with respect to the knowledge effect in the reading process. In addition, there has not been much L2 reading research that has combined quantitative and qualitative methodologies. As a result, this study aims to explore the roles of background knowledge and L2 grammar knowledge in L2 reading in terms of the strength of their effect and the and ways it takes place. To do that, the study combines quantitative and qualitative approaches. This chapter provides introduction to the study by first discussing the background and context followed by an outline of the research problem, aims, objectives, and research questions. Finally, this chapter suggests the significance of the study and concludes by providing the thesis structure.

1.2 Background to the study

Tertiary education settings typically require literacy in foreign languages, in particular English as a second language (ESL). Efficient L2 reading is strongly linked to students' academic performance and is considered a survival skill at the university level (Gunderson et al., 2020; Koda, 2020). This is one of the reasons it has attracted a lot of attention both theoretically and empirically.

Reading has been defined as "a process of constructing, knowledge integration, and building meaning" (Ruddell et al., 2019, p. 213). Comprehension, as the aim of reading, involves a complex interaction of reading processes and readers' knowledge and skills (Grabe, 2000; Grabe & Stoller, 2013). Lower-level reading processes comprise decoding, word-recognition, and parsing, whereas higher-level processes entail semantic and pragmatic processing at the local and global levels (Grabe, 2009). Readers use cognitive and linguistic sources of knowledge to build their own mental model of the text, which contains the most important propositions of the text that readers integrate with their background knowledge. This model has been referred to as the "situation" model of the text (Kintsch, 1998), and effective comprehension relies on how successfully readers integrate information from the text with their prior knowledge (Koda, 2005; Grabe, 2009). The reading process is mediated by cognitive mechanisms, including the working memory (Shin et al., 2019) and long-term memory, in which background knowledge is stored. Kendeou and O'Brien (2016, p. 159) contend: "Comprehension during reading is always subject to the interaction between individual differences in prior knowledge and the text." Consequently, they suggest that research needs to explain how individual differences in prior knowledge influence reading process and outcomes.

In their texts, writers make assumptions about what readers know. This is especially true in academic writing in disciplines where writers tend to expect the reader to have a grasp of the discipline knowledge and related specific terminology. Kintsch (1998, p. 313) emphasized that writers always rely on readers' knowledge and readers' text comprehension always depends on how they apply various types of knowledge. Moreover, background knowledge was found to be the driver and filter of the reading process (Alderson & Urquhart, 1999, p. 114). This suggests that readers engage in reading to build knowledge, and they construct text interpretation depending on their knowledge. Specifically, in the academic and professional contexts, readers have a need to read subject-specific texts to build knowledge and learn (Grabe & Stoller, 2019). By leveraging their knowledge, readers understand, interpret, and evaluate the texts they read. To the texts they read, readers bring various sources of knowledge, linguistic and non-linguistic, both of which affect the processes and products of their reading comprehension. Grabe (2009, p. 74) divides non-linguistic knowledge, also referred to as background knowledge, into four subcategories: a) general knowledge of the world, b) cultural knowledge, c) topical knowledge, and d) specialist expertise knowledge (also termed discipline or domain knowledge). He underscores that if discipline specialists read a text in their field, they are likely to read it differently than non-specialists with respect to the connections, inferences, and judgements they come up with. For this reason, discipline knowledge is crucial in academic and disciplinary reading (Shanahan, 2009, 2017). Because it interacts with other concepts and skills, it is not an easily studied factor (Grabe, 2009).

The design of empirical reading studies largely stems from the fact that reading is an invisible cognitive process that can only be approached in an indirect way by observing the way readers read or by examining what readers have gained through reading. In other words, the research focus is either on the process or the product of reading. Outcomes or products of reading have been typically investigated in quantitative studies that tend to operationally define reading comprehension as a score on a reading comprehension test. Process approaches to reading investigate observable features of reading, such as the readers' eye-fixations or pauses, assuming that these indicate various aspects of the reading process taking place. Nevertheless, there are clearly limitations as to what one can infer about reading processes on the basis of observable features of silent reading. Introspective (think-aloud) and retrospective methodologies have been developed and refined (Ericsson & Simon, 1984; Pressley & Afflerbach, 1995; Bowles, 2010; Israel, 2015) to deal with the "unobservable" elements of reading comprehension, especially higher-level processes.

The results of empirical studies exploring background knowledge are affected by how background knowledge is operationalized, so it is worth noting that the majority of studies so far have used indirect background knowledge measures. More specifically, background knowledge has mostly been measured through readers' self-assessment of content familiarity (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Hammadou, 2000; Lin, 2002; Brantmeier, 2003; Salmani-Nodoushan, 2003; Pulido, 2007; Leeser, 2007; Eidswick, 2010; McNeil, 2011; Lahuerta Martinez, 2013; Horiba & Fukaya, 2015; Shin et al., 2019) or dichotomously through their study discipline (Alderson & Urquhart, 1983; Koh, 1985; Peretz & Shoham, 1990; Ja'far, 1992; Clapham, 1996; Lee, 2007; Hill & Liu, 2012; Horiba & Fukaya, 2015). Only a handful of studies to date have used direct measures by testing knowledge (Uso-Juan, 2006; Rydland et al., 2012; Lin & Chern, 2014; Karimi; 2017; Hwang & Duke, 2020; Song & Reynolds, 2022). It is worth noting here that self-reporting may be prone to subjectivity and people may possess certain knowledge even without a formal discipline background. Thus, the assumption about one's knowledge based on their study discipline may only be a rough measure of the variable. Likewise, people outside a given discipline may possess particular discipline-related knowledge because of their interest or experience. It also cannot be assumed that all people within a certain discipline have the same amount of knowledge. In addition, it should be highlighted that especially in studies dealing with discipline-related knowledge, the texts used for reading comprehension need to be assessed for specificity; otherwise, the background knowledge effect may not be detected (Clapham, 1996).

1.3 Research problem

L2 reading has been defined as a multifaceted cognitive process that involves the interaction of language and knowledge-based sources (Grabe, 2009; McNeil, 2012); moreover, "the interaction of variables is even more pronounced in L2 reading" (Uso-Juan, 2006). Drawing on the findings and gaps in existing reading research, three key problems have been identified and are addressed in this study.

Firstly, numerous studies have investigated the effect of background knowledge in conjunction with L2 proficiency. However, these studies have traditionally used indirect operationalizations of background knowledge in the form of assumed knowledge either as readers' study discipline or readers' self-reporting of their topic familiarity and knowledge, which has been identified as a potential reason for inconsistencies in reading studies (Cervetti & Wright, 2020). This suggests a need for a comparison of direct and indirect operationalizations of knowledge in one study.

A second issue is that despite ample L2 research, studies involving mixed methods are scarce, and there is still a need to explore how prior knowledge affects the reading process (Karimi, 2016; Nassaji, 2007). Most L2 reading research has focused on the importance of variables in the reading process, but not on how they operate. As a result, a number of studies have suggested the need for process studies (Yamashita, 2002; Koda, 2012), introspective methods (Uso-Juan, 2006), and think-aloud methods

(Pritchard, 2006; Horiba & Fukaya, 2015). Such studies would shed more light on the allocation of cognitive resources during text processing.

The third issue addressed here is that within a wide array of background knowledge studies, there is still scant research into discipline-related knowledge (McNeil, 2012; Hwang, 2019). Hwang (2019, p. 9) also argues a case for more studies addressing the relationship between domain knowledge and reading comprehension across different countries.

1.4 Research aim, objectives, and questions

Given the lack of mixed research and scant investigations of different operationalizations of discipline-related background knowledge in L2 reading, this study aims to address this gap by using mixed methods and two operationalizations of knowledge. It aims to explore the relationship between discipline-related background knowledge and grammar knowledge in L2 reading of subject-specific texts, by determining and evaluating the strength of this relationship as well as the ways it operates. In addition, the study aims to investigate the distinction between tested and self-reported knowledge.

Guided by the main research aims, the study pursues the following six specific research objectives:

1. To identify the contribution of background knowledge in L2 reading comprehension with regard to its strength, interaction with L2 grammar knowledge, and the ways it takes place.
2. To compare the contributions of background knowledge and L2 grammar knowledge in L2 reading comprehension.
3. To identify the existence of any L2 threshold necessary for the effect of background knowledge to take place.
4. To test the existence of compensation between background knowledge and L2 knowledge.
5. To compare the contribution of tested and self-reported background knowledge in L2 reading by using two different operationalizations of background

knowledge: a) a test of text-related background knowledge and b) readers' self-reports of their familiarity with the text topic.

6. To detect the ways readers use background knowledge in L2 reading.

Based on these research aims and objectives, research questions 1–6 are formulated at the end of Chapter 2, whereas research questions 7 is presented at the end of Chapter 3.

1.5 Significance of the study

This study aims to contribute to the body of knowledge in L2 reading by determining and evaluating the relative role of discipline-related background knowledge and L2 grammar knowledge in L2 reading. It addresses the lack of mixed research by combining the quantitative and qualitative approaches. In addition, it responds to the lack of studies comparing the difference between the role of tested and self-reported background knowledge. In this way, this research aims to add value to a better understanding of the L2 reading process and add insights to L2 reading instruction and research.

1.6 Structure of the thesis

In Chapter 1, the context of the study is introduced. The research problem, aim, and objectives are identified, and the value of such research evaluated. In Chapter 2, the existing theoretical and empirical literature is reviewed to identify key findings in L2 reading research with respect to the role of background knowledge and grammar knowledge in the reading process. The research gaps are discussed. Chapter 3 provides a survey of literature on think-aloud methodology and justifies its use in this research. Chapter 4 presents the framework of the study. It justifies the adoption of both quantitative and qualitative approaches and discusses the research design, including its limitations. Chapter 5 gives a detailed presentation of the analyses and results in the testing study. Chapter 6 presents the results of the think-aloud study. Chapter 7 discusses the results from both quantitative and qualitative studies by comparing them with the existing research in L2 reading, both theoretical and empirical. It explains and interprets the results by highlighting how the findings fit into the context of existing research. It underscores what is new and evaluates the significant findings. Chapter 8 gives a summary of the study, its contributions, and implications for L2 instruction and reading research. It also identifies the limitations of the study and links them to suggestions for future research.

CHAPTER 2
BACKGROUND KNOWLEDGE AND READING COMPREHENSION

2.1 Chapter aim and overview

This chapter reviews the theoretical and empirical research on the role of background knowledge in reading comprehension. Knowledge is a reader-based, text-based, and task-based variable, so the chapter addresses three clusters of factors that relate to the reader, the text, and the task. I first look into the rationale for including background knowledge in reading comprehension studies by exploring the theoretic underpinning provided by different reading models. I then examine different types of background knowledge and survey empirical studies of L2 reading. Specifically, I review L2 reading studies with regard to the effect of background knowledge they have detected and the relationship between background knowledge and L2 proficiency. After reviewing topic and domain knowledge studies, I discuss open issues and then state my research questions based on the research gap.

2.2 L2 reading research

Learning L2 for academic and professional purposes as well as keeping abreast with the development in a number of disciplines requires reading in L2. Effective reading in L2 is central to academic and professional development, so it is not surprising that reading has generated sustained interest among cognitive psychologists, linguists, and teaching practitioners. The last fifty years have seen the publication of a number of volumes on reading, the proposal and refinement of several models of reading, and the publication of numerous empirical studies investigating various aspects of reading: skills that underlie reading, reading comprehension processes, teaching and testing of reading. Process and product-oriented research was based on qualitative and quantitative methodologies, resulting in a wide range of empirical studies from correlational to introspective.

Though slightly different in their focus, L1 and L2 reading research have both tried to account for individual differences in reading, focusing on variables related to either the reader, the text, or the task. In L2 reading, all three groups of variables are affected by the specific reading context. First, readers bring to reading not only different language knowledge but also different knowledge of the world, including their particular discipline. Second, the amount of knowledge assumed by texts may be crucial for readers' comprehension of a specific text. Finally, when texts are read for information and to fulfil different study or professional tasks, the style of reading will vary accordingly. Background knowledge and how this knowledge is used when reading may contribute to effective reading comprehension in academic and professional contexts. The effect of background knowledge has been mostly examined in product-oriented studies of reading comprehension, but there have not been as many process studies investigating how background knowledge is deployed when reading. An obvious reason for the lack of such studies may be the notorious difficulty of tapping into a process as elusive and difficult to observe as reading.

2.3 Theoretical underpinning of background knowledge as a variable in reading comprehension

Even though L2 reading studies in many ways reflect concerns voiced in L1 reading, the nature of L2 learning and teaching has led to a slightly different angle of L2 reading research, with linguistic factors gaining more saliency. Factors influencing reading have been first formulated as a juxtaposition of linguistic skills vs reading skills in Alderson's question (1984) whether reading is a language problem or a reading problem. As a result, language concerns motivated research interest in the contribution of L2 proficiency in reading comprehension. The research angle was skewed towards linguistic and learning-teaching aspects with a focus on skills components that underlie reading and may presumably be taught, either on the linguistic or processing side.

Theoretical models that try to conceptualize L2 reading were proposed by Coady (1979), Bernhardt (1991b), and Hoover and Tunmer (1993). Labelled as componential, these models do not attempt to give a sequential picture of the reading process, but they examine the skills (i.e. components of ability) that underlie reading comprehension in L2. These models differ in the number of components they identify and the relative stress they place on them. They also stem from different contexts, either purely theoretical (Hoover & Tunmer, 1993; Coady 1979), or empirical (Bernhardt, 1991b). However, the value of componential models is that they offer a conceptualization of the skills that underlie the reading process and can be empirically tested. Overall, componential approaches to L2 reading seek to determine variables that lead to reading ability differences and inform instruction (Jeon & Yamashita, 2020).

2.3.1 Two-component models of L2 reading

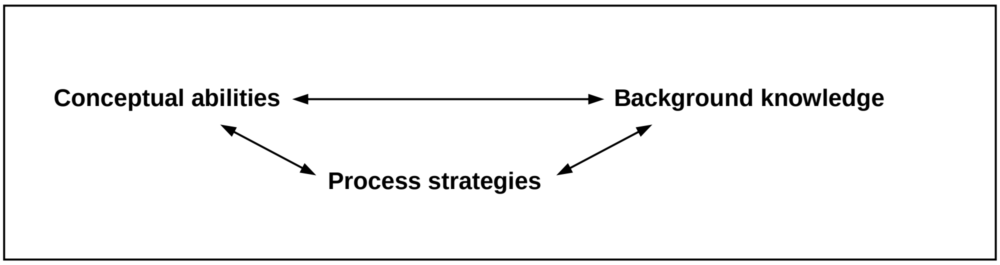
The Hoover and Tunmer model (1993) offers a "simple view" of reading regarding the number of components, pointing merely to two reading components: first, decoding in the sense of word recognition, and second, linguistic comprehension. The first component is understood as matching written graphical input indirectly – that is, via the phonological route – to the readers' mental lexicon. On the other hand, linguistic comprehension entails readers' interpretation of the lexis accessed at a word, sentence, or text level. The inclusion of components in this model has a fairly common-sense interpretation. If one possesses a command of L2, but cannot decode, the result will be failure to comprehend; the same is true if one can decode but cannot speak the L2. Supporting evidence for this two-way model comes from language or reading impairment studies: dyslexics are linguistically competent but restricted in decoding; conversely, hyperlexics are highly efficient at decoding, but they have problems in correct language use (Urquhart & Weir, 1998).

Although the selection of both components in this model cannot be contested, it appears that there is more to the skills underlying reading than just decoding and language. This can be illustrated by taking an example of a highly specialized or ambiguous L2 text in which ideas are not expressed overtly or they are decontextualized. Even though readers may possess both the skills from the Hoover and Tunmer model, their reading effort may result in at least a partial failure. It is for these reasons that one needs to turn to the models that encompass experiential and cognitive elements that go beyond merely decoding and linguistic skills.

2.3.2 Three-component models of L2 reading

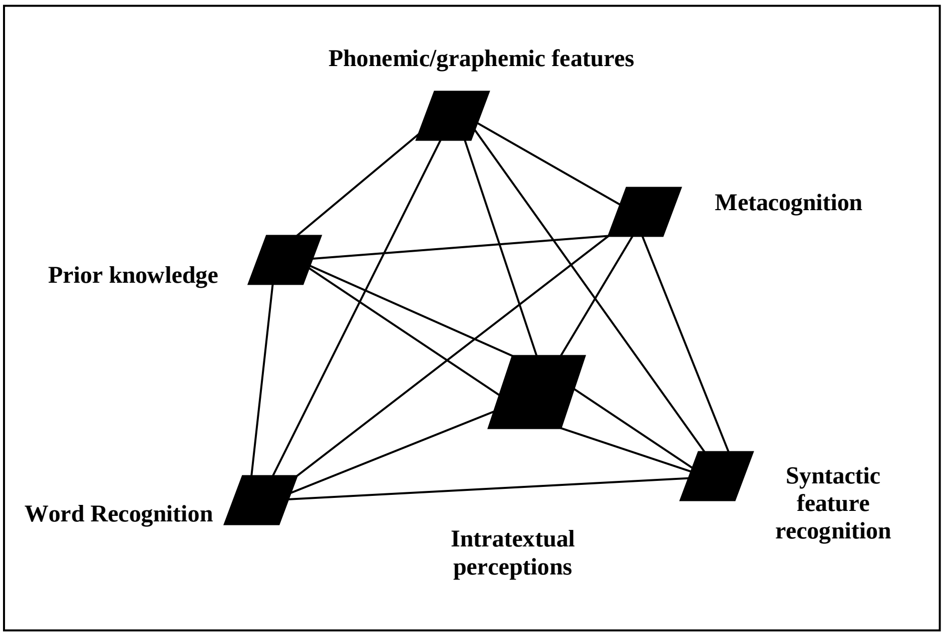
The two most influential three-component models of L2 reading (Coady, 1979; Bernhardt, 1991) both involve a variable referring to non-linguistic knowledge and experience the reader brings to reading. The three interacting skills in the Coady model (Figure 2.1) are conceptual abilities, process strategies, and background knowledge. Conceptual abilities stand for the reader’s intellectual capacity, whereas process strategies refer to matching of graphemes, phonemes, and morphophonemes, syllables and morphemes, and knowledge of syntax, lexical meaning, and contextual meaning. The third element of Coady’s model and a contributing variable to reader comprehension is background knowledge. Coady interpreted his model as flexible, thereby accounting for a different use of processing strategies, either as a developmental feature in the reader or as a feature resulting from different L2 knowledge.

Figure 2.1 The Coady Model of L2 Reading Process (Coady, 1979 in Bernhardt, 2011, p. 24)



Similar to Coady, Bernhardt included a component called prior knowledge in her 1986 model (Figure 2.2), and background knowledge in her 1991 model (Figure 2.3). The Bernhardt 1986 constructivist model of L2 reading (Figure 2.2) is an elaboration of the Coady model with interacting components, both intratextual and extratextual: phonemic/graphemic features, metacognition, perceptions, word recognition, and prior knowledge. These components are described as interacting in the process of the reader’s formation of a mental representation of the text. The model is described as “interactive and circular” and allows for interactions among all components at any point (Bernhardt, 1986).

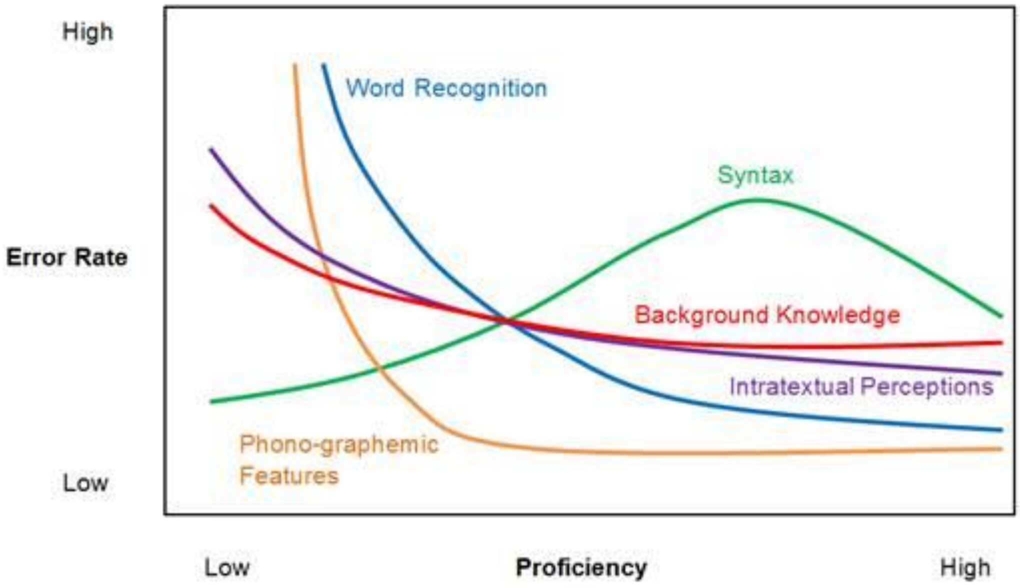
Figure 2.2 The Constructivist Model of L2 Reading (Bernhardt, 1986 in Bernhardt, 2011, p. 30)



The Bernhardt 1991 model (Figure 2.3) provides more detail and was derived from an empirical database of L2 reading rather than from theoretical assumptions. The model was constructed as a combination of five graphs representing five reading factors which are plotted in a coordinate system, with levels of proficiency shown on the x-axis and the error rate on the y-axis. The graphically presented reading factors are phono/graphemic features, word recognition, syntax, background knowledge, and intratextual perceptions. This model contrasts with prior models in several ways: it brings together reading factors in a developmental picture depicting error rate as a function of proficiency level (Bernhardt, 1991, pp. 168–71). The basic tenets of the model are as follows:

1. Text processing abilities develop over time and are not learned as unitary entities.
2. Reading factors do not replace each other in a linear fashion, but interact among themselves.
3. Errors in reading comprehension show one's development in literacy.
4. There are common features in L2 text processing among different readers as well as different L2's.
5. No definite terminus is assumed, so no reader can ever be 100% proficient and have a 0% error rate.

Figure 2.3 The Theoretical Distribution of Reading Factors in L2 (Bernhardt, 1991, p. 169)



The three language-based factors in the Berhardt 1991 model (word recognition, phono-graphemic features, and syntactic feature recognition) show great developmental change, with phono-graphemic and word recognition errors exponentially falling, and a Gaussian shaped curve for syntax errors. At advanced proficiency levels the error rate plummets yet remains comparatively higher than errors related to all other factors. The curves depicting the knowledge-driven factors (background knowledge and intratextual perceptions) are exponential but relatively more constant than any other factor curve. The error rate concerning background knowledge is lower than word recognition and phono-graphemic features at the beginner stage. At intermediate proficiency, word recognition and phono-graphemic errors drop sharply, whereas background knowledge and intratextual errors decrease gradually and then level out at high proficiency. At high proficiency levels, the highest rate of errors are those to do with syntax. According to Bernhardt (1991, p. 170), the comparison of the background knowledge curve with all other curves shows:

> "the interaction of knowledge and language as proficiency develops. In initial stages, whatever knowledge a reader may bring to a task, may override linguistic knowledge; hence there are relatively more knowledge-based errors than syntax errors. As a reader's linguistic knowledge grows, however, it begins to override knowledge-driven inferencing. In other words, a reader begins to rely more on the *language* and less on what he/she *thinks* the language contains."

The strength of Bernhardt's 1991 model can be ascribed to the fact that it was derived from a L2 reading database rather than purely theoretical premises, and the assumptions of the model can be validated by replication.

2.3.3 Linear models of L1 reading

Linear models belong to process models of L1 reading and they unpack the process of reading as a linear series of sequential phases, divided into low-level and high-level processing. According to their prevailing emphasis on the type of processes involved in reading comprehension, linear models can be divided into bottom-up and top-down.

Bottom-up models, also called 'text-driven' models (Gough, 1972), are derived from behaviourist positions, and they stress a stimulus-and-response type approach. They posit that visual decoding lies at the heart of reading, with the reader proceeding in a linear fashion from letters to words, sentences, and texts. More specifically, reading is understood to involve a hierarchy of five processes starting with eye-fixation, followed by absorption of the visual stimulus, letter identification, phonological representation and, finally, understanding. The role of the reader in reading is that of a text-decoder and absorber of the written input, striving to reconstruct the author's meaning.

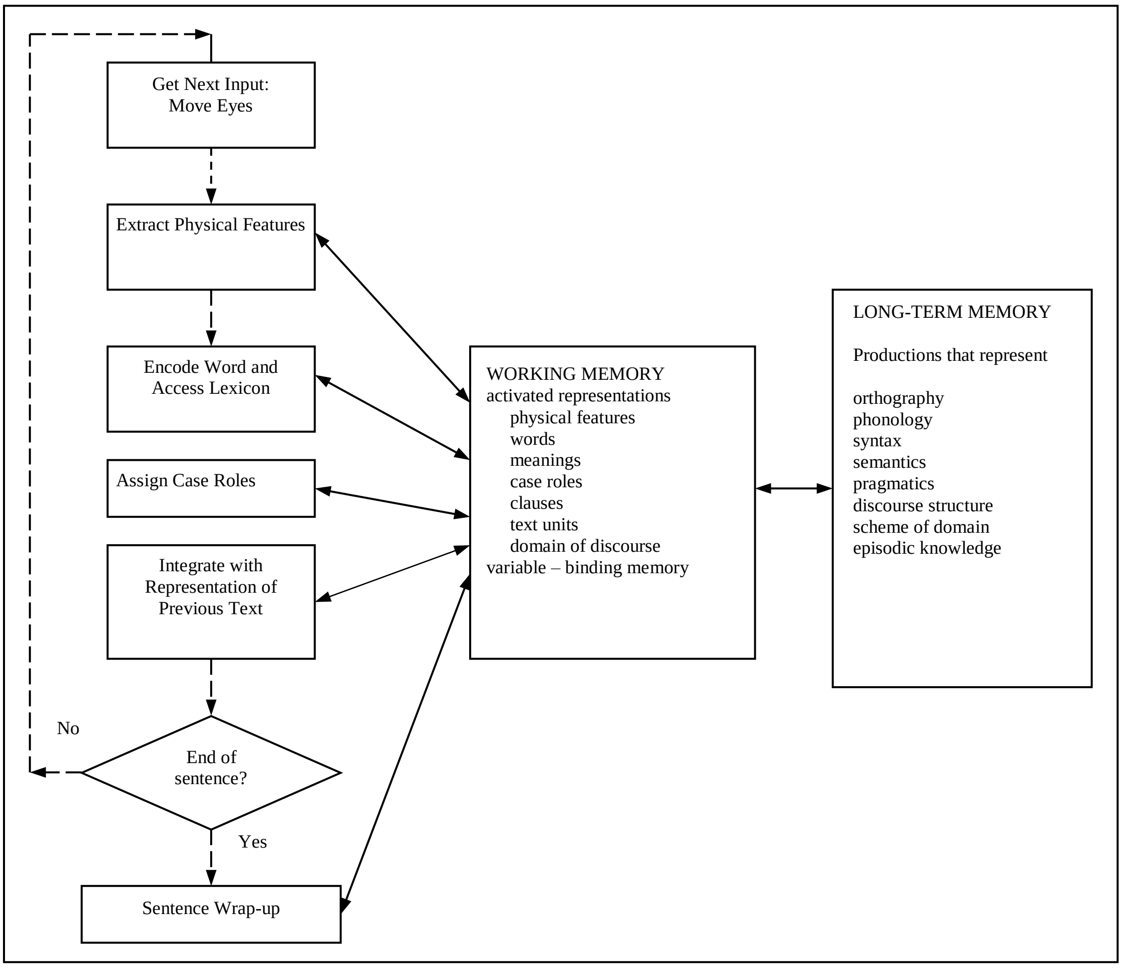
Conversely, top-down models (Goodman, 1967) take as the starting point readers with their expectations and purposes of reading. They focus mainly on the knowledge and skills that readers bring to texts, and thus are called 'concept-driven' models. These models shift the focus to higher-level processes and offer a sequential picture from the top-down perspective. The reading process begins with readers' hypotheses which may or may not be confirmed. According to the top-down view, there is no text on its own, rather there are different text representations that different readers make when reading a text. In Goodman's terms, reading is a "psycholinguistic guessing game" (1967), whereby reading is a process of readers' forming, testing, confirming, or refuting their hypotheses. While bottom-up models do not account for higher-level processes, this cannot be said of top-down models. They stress that lower-level skills need to be automatized (Goodman, 1967), so that more time can be spent on higher-level operations such as: inferencing, predicting, interpreting, and integrating. Good readers have quick and automatized lower-level skills, which enables them to pay more attention to meaning and text representation. Although bottom-up and top-down models share a linear approach to reading, they offer a picture of the reading processes from two different perspectives.

2.3.4 Interactive models of L1 reading

Another position in reading research is in favour of interactive models, which posit that the processing involved in reading may not be only at one level, either lower or higher, but rather a combination of the two, which is crucial to text comprehension (Rumelhart, 1977; Grabe, 1991). Reading involves using various decoding and word recognition skills as well as a variety of comprehension and interpretation skills. Knowledge involved in the reading process is further defined. Rumelhart (1977) suggests that semantic, lexical, orthographic, and syntactic information from the text interact with each other in the pattern synthesiser. The models devised by Just and Carpenter (1980) and Rayner et al. (2011) focus in particular on the interaction between processes at different levels.

The interactive nature of the model was further elaborated in Stanovich's compensatory model (1980), positing that the reader may accommodate deficiencies or difficulties in reading by resorting to a different level of processing. So, for instance, poor language proficiency may be to some extent compensated for by superior background knowledge. Even though this interactive model was devised for the L1 situation, it can also be applied to the L2 context. Interactive models clearly acknowledge the role of background knowledge in the reading process. The Just and Carpenter model (1980) in Figure 2.4 identifies the "scheme of domain" and "episodic knowledge", whereas the Rayner et al. model (2011) refers to "real world knowledge". In both models, this knowledge is stored in long-term memory, but interacts in the reading process through working memory.

Figure 2.4 The Just and Carpenter Model of Reading (Just and Carpenter, 1980, p. 331)



2.3.5 Schema theory

Research into knowledge in reading comprehension has evolved from schema theory in cognitive psychology (Bartlett, 1932). This has been an influential approach that aimed to account for the pattern of human remembering and forgetting (Rumelhart, 1975; Schank, 1975; Schank & Abelson, 1977). The main tenet of the theory is that what we remember is affected by what we know; thus, a person's experience and knowledge play a crucial role in accommodating new information. The theory also posits that knowledge stored in our memory is organized in a hierarchy of schemata, with a schema as a basic unit of one's memory, defined as

> "a simplified, generalized, mental representation of everything we understand by a given type of object or event, based on the past experience" (Gross, 1996, p. 300).

Rumelhart and Norman (1983) defined five major characteristics of schemata related to their structure and representation (quoted in Gross, 1996). First, schemata contain slots; some of them are fixed, meaning that we always associate a particular event or object with that schema. However, slots in a schema may also be open and are specified for each individual situation. To illustrate the point, Shank and Abelson (1977) refer to a restaurant schema which may contain fixed objects, activities, and roles, such as tables, food, the waiter, the cook, eating, and paying; whereas a particular meal may figure as an open slot: the food may be delicious or not, and so on. Schemata thus form a person's generic knowledge and can be related to specific situations. Second, schemata form systems, which may be a part of a larger or different schemata system (e.g. eating in a restaurant as part of entertainment). Third, schemata represent both concrete objects, events, and abstract concepts. Fourth, schemata represent knowledge rather than definitions. Finally, schemata are active recognition devices which enable us to interpret new, ambiguous, or not sufficiently defined information in line with the knowledge we have. This implies that in reading comprehension, schemata play a key role in readers' filling in the gaps in the text, predicting, and evaluating.

Schema theory has not only exerted a strong influence in cognitive psychology and reading research, but has also drawn a lot of criticism. The theory was criticized for being "unprincipled" (Eysenck & Keane, 2005, p. 268) as schemata are not defined in sufficient detail and were found to be both "too definite and too sketchy" (Bartlett, 1991 in Urquhart & Weir, 1998, p. 70). Several problems were identified: schema theory cannot account for the formulation of basic schemata, which are formed when one does not have any knowledge to interpret new information. Moreover, schema theory fails to explain the fact that we may remember new, unusual things or events much better than the old usual ones. Rayner et al. (2011) pinpointed a key problem of schema theory to be the size of schemata and the definition of schemata as basic units of knowledge. They argued that because it is difficult to determine the size of schemata, they are interpreted to be "nested", implying various sizes, thus inappropriate to be called "units" of real-world knowledge. They find this to be the reason a number of schema theorists argue that there is a limited number of generic schemas, which may incorporate individual, specific schemas.

Rayner et al. (2011) critically review schema theory at two levels: the episode and story level. Regarding the former, they refer to studies claiming that all information that cannot fit a schema is lost, and information is only encoded in terms of general schemas. They argue that memory for information that cannot fit the existing schemas may be worse (e.g. random sentences vs a coherent text), nevertheless they suggest that readers use some sort of schemas even when trying to remember random sentences. Graesser et al. (1979) found supporting evidence for the schema view of text comprehension. When requested to single out statements that were used in a passage, the subjects remembered atypical statements better than typical ones. Rayner et al. (2011) point to a paradox of schema theories, which postulate that something is not well-remembered if it doesn't fit into a schema. In addition, they claim that atypical information is remembered better than typical. Typicality is found to both help and hinder the memory, yet it is not clear when and how this occurs. In addition, Rayner et al. (2011) argue that schema networks cannot represent important aspects of meaning. They quote two strands of research that identify shortcomings of schema theory: studies of causality (Trabasso & Sperry, 1985) and studies of mental models (Johnson-Laird, 1983). First, schemas do not seem to explain "whether the text makes sense" and they seem to treat in the same way the texts with or without causal or sequential links. Counter evidence (Trabasso & Sperry, 1985) suggests that texts with causal linking are remembered better than those without. The rated importance of a statement depends on both the connections and causal relation between sentences. By causal analysis they predicted the immediate and delayed recall as well as the time to retrieve information. Overall, Rayner et al. (2011) contend that by attempting to explain how information is stored in memory, schema research has accounted for several important points: first, memory of texts that make sense is better than that of those that do not. Second, there are certain units in the text, called episodes, which are important in remembering. Finally, redundant information appears not to be stored well in memory. They agree that some memory tends to be stored in schemas, but express doubt that all memory is encoded in this manner. They argue the existence of a dual storage mechanism (Kintsch & van Dijk, 1978), or parallel storage of two types of memory: literal and gist memory. As supporting evidence, they refer to larger-scale schema experiments or "story grammars" which failed to prove schemas as the only form of memory representation (Rumelhart, 1975; Thorndyke, 1977; Mandler & Johnson, 1977).

Even though schema theory has been questioned by many researchers on the grounds that it cannot be explicitly defined, several authors have pointed to the value of the theory and areas of its influence. Grabe (1991, p. 389) acknowledged that

> "in spite of the fact that schema theory is not a well-defined framework for the mental representation of knowledge, it has been an extremely useful notion for describing how prior knowledge is integrated in memory and used in high-level comprehension processes."

Referring to L2 reading research, Grabe (2009, p. 80) argues that "the notion of schemata remains a useful metaphorical explanation for many experimental results", which has particular relevance for the teaching of reading as indicated in schema activation studies (Carrell, 1984; 1987; Carrell & Eisterhold, 1983). These studies have shown that activation of content knowledge crucially affects reading comprehension and text recall, which has important pedagogic implications for L2 reading. If the reader activates their background knowledge, or is provided with the knowledge relevant to the text, this is likely to facilitate text interpretation, especially when the text is less overt and there are a lot of gaps to be filled by the reader. The value of providing background knowledge is in enabling not only text comprehension but also readers' evaluation and personal response to the text (Nuttal, 1996). It has also been suggested that L2 difficulties could be mediated by building readers' background knowledge (Wallace, 1992). Activating background knowledge in pre-reading activities may indeed be relevant to the L2 reading situation, yet it had better not be undertaken solely as a compensatory strategy for poor decoding automaticity (Paran, 1996).

In summary, even though intuitively appealing for its attempt to find the structure of how knowledge is stored and retrieved, schema theory leaves many open issues and has limited usefulness in reading research (Sadoski et al., 1991). There are four major arguments constraining the value of schemata for reading research. First, a schema is defined as both fixed and adaptable to new situations, which makes it difficult to use in practice. Second, on terminological grounds: if the term *schemata* is no more specific than *background knowledge* but is used interchangeably, why use it? Third, as a notion, schemata lack a detailed definition and experimental supporting evidence. Fourth, the activation or acquisition of schemata referred to by L2 researchers is loosely defined and schemas seem to be activated by readers "at the drop of a hat". Therefore, some researchers prefer to use the term *background knowledge* rather than *schema* (Urquhart & Weir, 1998).

2.3.6 Construction-Integration (C-I) models

The dominant current models of reading comprehension are the Construction-Integration (C-I) models (Kintsch, 1998; Perfetti, 1999). Pearson and Cervetti (2017) believe these models "provide the best explanation of how readers make meaning from the printed word." (2017, p. 41) and have thus become "the dominant paradigm in explaining the conceptualizations of both basic processes and pedagogical practices for comprehension" (2017, p. 43). Construction-Integration models posit that reading comprehension encompasses at least three levels of processing. The first level or "surface structure" refers to the linguistic structure of the text. It involves readers' decoding of words and phrases into a string of propositions, which is mainly bottom-up. The second level or "textbase" is a product of readers' encoding of semantic and rhetorical elements of a text. This is a top-down process requiring the link to the long-term memory. The third processing layer or "situation model" is readers' mental representation of the text. It is a product of readers' integrating the text with their background knowledge or information in readers' long-term memory. When reading, readers make inferences and elaborate ideas from the text. If text propositions and inferences align, readers can build a coherent mental representation of the text. Cervetti and Wright (2020, pp. 238–9) stress that constructing the situation model means building a reciprocal relationship between reading comprehension and knowledge. Knowledge supports comprehension and comprehension refines and builds new knowledge. The augmented knowledge is stored in long-term memory as prior knowledge and is used in subsequent reading, both within the same text and in new texts.

In summary, models of reading offer theoretical support for the inclusion of background knowledge as a variable in reading comprehension. Background knowledge is a factor in componential models (Coady, 1979) that identify the skills that underlie reading, interactive models (Goodman, 1967; Smith, 1971) that highlight higher-level processes, and Construction-Integration models (Kintsch, 1998; Perfetti,

1999) that include the third processing layer during which readers integrate the text with their knowledge. Interactive models also posit an intriguing stipulation that background knowledge may compensate for reader's linguistic deficiencies. Notably, these models assign to the reader the role of an active agent who determines the reading process by what s/he brings to it in terms of linguistic and reading skills as well as background knowledge.

2.4 Empirical studies of L2 reading

A major impetus behind numerous L2 empirical studies is the aim to locate the sources of difficulties that readers encounter when reading texts in L2. This motivation meant that studies looked into three groups of factors that refer to the reader, text and task (Table 2.1). First, reader-bound variables include L2 proficiency, background knowledge, processing, and cognitive abilities that readers bring to texts, as well as affective factors that include readers' purposes for reading, expectations, motivation, and interest. Second, text-bound variables can be linguistic and non-linguistic. Linguistic text elements comprise lexis, syntax, and rhetorical structure, whereas non-linguistic elements refer to the text topic as well as the knowledge the writer assumes the reader to have. Third, task-bound variables are related to methods of testing reading comprehension (e.g. multiple-choice questions, cloze tests, short question answers, recall, summaries) that may affect reading comprehension. It should be noted that distinctions among these elements are not completely neat and clear cut. Syntax is seen as a function of rhetorical structure of text, but it involves arrangement of vocabulary. Vocabulary is used to express concepts and ideas of the writer, while it is an integral part of readers' prior knowledge and experience.

Table 2.1 Factors accounting for differences in L2 reading

READER-BOUND FACTORS	**Linguistic proficiency**: vocabulary, syntax, rhetorical structure **Background knowledge**: subject matter, discipline, the world, culture **Processing and cognitive abilities**: low-level (e.g. decoding), high-level (e.g. inferencing) **Affect**: motivation, interest, purposes, and expectations
TEXT-BOUND FACTORS	**Linguistic:** vocabulary, syntax, rhetorical structure, genre, readability, explicitness **Non-linguistic**: topic and content, assumed background knowledge, specificity
TASK-BOUND FACTORS	**Task type**: multiple-choice questions, short-answer questions, cloze, recall, summary

Empirical research in L2 reading examines factors related to readers, texts, and tasks. It can be divided into three strands roughly parallel to the three components featuring in the Coady model (1979). The first one focuses on the differences between reading in L1 and L2. The second group of studies is concerned with the existence of the L2 threshold level beyond which skills transfer is possible. This group includes a number of studies that compare good and poor readers and aims to single out abilities which are critical for good L2 reading and reading instruction. Finally, background knowledge studies examine the influence of different types of prior knowledge that readers bring to reading. In the following sections, I review those L2 empirical studies that are of closest relevance to the aims of the present study.

2.4.1 Defining background knowledge

Prior to reviewing empirical studies of L2 reading, I will look into how background knowledge has been defined in reading research. A number of studies have attempted to determine elements of background knowledge (Carrell, 1983, 1987; Bernhardt, 1991; Urquhart & Weir, 1998). Carrell (1983) identified three background knowledge factors: context, transparency, and familiarity. *Context* corresponds to the content knowledge gathered by the reader from the headlines or pictures accompanying the text; *transparency* refers to the extent the textual elements disclose the meaning of the text; and *familiarity* denotes reader's specific knowledge of the content area. As the first two elements are clearly more text-bound, but may affect reading comprehension in interaction with the last element, it is familiarity that is closest to a general definition of the background knowledge.

Focusing on the familiarity element in the Carrell classification (1983), Bernhardt (1991) distinguishes between three types of background knowledge: local-level knowledge, domain-specific knowledge, and culture-specific knowledge. *Local-level knowledge* is defined as highly individualized, idiosyncratic knowledge a person possesses or implicit knowledge particular groups may have. *Domain-specific knowledge* relates to one's educational, professional, and work background, as well as additional knowledge one acquires by self-study and interest in a particular area. It may include not only declarative knowledge about things, but also procedural knowledge about how to do things. As the third type, Bernhardt (1991) lists *culture-specific knowledge*, dividing it into ritualistic and cultural historic knowledge. Members of a specific culture acquire implicit ritualistic knowledge, which is passed from one generation to another, referring to family celebrations, national holidays, religious ceremonies, etiquette, values and so forth. In line with her classification of the types of background knowledge, Bernhard (1991) lists three groups of studies investigating the effect of background knowledge: first, studies dealing with subject or topic knowledge; second, studies looking into cultural background knowledge; and finally, studies inquiring into the activation of background knowledge. However, there may be a certain degree of overlap between the categories mentioned. According to Widdowson (1979), academic domains may also be interpreted as cultures formed by different discourse communities, like cultural groups formed by people sharing ethnicity, history, religion etc. Thus, subject-specific knowledge can also be interpreted as cultural knowledge, both being variations of content or topic background knowledge. Due to this similarity, Urquhart and Weir (1998) propose that content knowledge and cultural knowledge should be considered as a whole labelled *background knowledge*. As discussed before, they also argue for avoiding the term *schema*, because they do not find it any more specific or meaningful than background knowledge.

Carrell (1984, 1987) makes a distinction between *formal* and *content schemata*, with the former referring to the rhetorical organization of a text, which may be more or less familiar to the reader. Bernhardt (1991) and Urquhart and Weir (1998) see formal schemata as part of a broader concept of *literacy,* defined by Bernhardt as the way one approaches texts, involving the reader's awareness of the purposes and results of reading. Urquhart and Weir add two more elements to the literacy component: 'cohesion' and 'text structure', quoting different positions on the matter, as some authors consider awareness of cohesion as a part of linguistic knowledge (Halliday & Hasan, 1976), whereas others see cohesion to be connecting text-knowledge and world-knowledge. As cohesion and text structure actually correspond to Carrell's formal schemata, it can be concluded that formal schemata may be seen as part of literacy. Carrell (1988) also referred to *linguistic schemata* or knowledge of the language. However, it should be noted that in the majority of L2 reading studies, language knowledge is considered as a variable separate from the rest of knowledge a person has, and is either operationalized as linguistic proficiency, or broken down to its components such as knowledge of syntax or vocabulary.

In reading research, the expressions *prior knowledge* and *background knowledge* are often used interchangeably. Some authors, however, make a distinction between the terms as regards the level of specificity by defining *prior knowledge* as "a general collection of facts, concepts and experiences" (Hennessy, 2021, p. 149), whereas *background knowledge* is defined as more specific to situations, problems, concepts, and information relevant to a particular text. Fogarty et al. (2021, pp. 97–9) make a distinction between prior knowledge, background knowledge, content-specific knowledge, and domain-specific knowledge. They relate each type of knowledge to three different tiers of vocabulary: tier one is everyday language, tier two general academic vocabulary, and tier three domain-specific vocabulary. *Prior knowledge* is typically expressed with tier one vocabulary, whereas *background knowledge* combines tier one and tier two language. *Content-specific knowledge* is rendered through tier two language, whereas *domain-specific knowledge* is based on tier three vocabulary that is not considered common knowledge.

In their comprehensive reviews of reading research, Grabe (2009), Kendeou and O'Brien (2016), and Cervetti and Wright (2020) distinguish between four types of knowledge: 1) world knowledge, 2) topic knowledge, 3) domain knowledge, and 4) cultural knowledge. *World knowledge*, also termed *general knowledge*, is the most general type of academic knowledge that may or may not be relevant to the content of the particular text. *Topic knowledge* or *subject-specific knowledge* is closely linked to the topic of a text. *Domain* or *discipline knowledge* refers to the grasp of a particular academic discipline or a field of study that the text topic relates to. *Cultural knowledge* encompasses the socio-cultural experiences of the reader. This classification of background knowledge is based on the breadth of knowledge and its proximity to the text content. World knowledge is the broadest in scope and not related to the text content, whereas topic knowledge is narrow, specific, and directly related to the text content. This terminological classification of background knowledge will be used in this study and is summarized in Table 2.2.

Table 2.2 Types of background knowledge in L2 reading studies

Type	Definition	References
World knowledge also: **General knowledge**	Wider academic knowledge unrelated to a particular text.	Koda, 2005; Grabe, 2009; Cervetti & Wright, 2020
Domain knowledge also: **discipline knowledge, specialist expert knowledge**	Knowledge linked to an academic discipline or broader study field or thought.	Alexander et al., 1991; Bernhardt, 1991; Koda, 2005; Grabe, 2009; Cervetti & Wright, 2020
Topic knowledge also: **specific knowledge**, **subject-specific knowledge**	Specific knowledge closely connected or relevant to a particular text topic. A subset of domain knowledge.	Grabe, 2009; Kendeou & O'Brien, 2016; Cervetti & Wright, 2020
Cultural knowledge also: **culture-specific knowledge**	Knowledge built by readers' socio-cultural experiences.	Koda, 2005; Bernhardt,1991; Grabe, 2009; Cervetti & Wright, 2020

2.4.2 Background knowledge and language proficiency in empirical studies of L2 reading

Theoretical and empirical studies provide ample evidence supporting the role of background knowledge in L1 and L2 reading. L2 reading, in particular, has generated a lot of interest and a large number of studies have been conducted with L2 readers. The reason for this may be twofold: L1 readers may rely more heavily on text-driven processes such as automatic word recognition, and they do not need to resort to compensation strategies, such as the use of top-down processes, including the deployment of background knowledge. On the other hand, L2 readers may be faced with at least two sources of difficulties: linguistic and reading. If they have poor language knowledge or poor reading skills, they may resort to using other knowledge, skills, and strategies they possess to overcome problems and aid their reading comprehension. In other words, due to language and reading deficiencies, L2 readers may use higher-level reading processes including background knowledge in order to substitute for the lower-level processes they lack.

Language proficiency features as a variable in most L2 studies exploring the role of background knowledge in reading comprehension. Two approaches to interpreting the relationship between language proficiency and background knowledge can be observed. The first one reflects the tenets of compensatory interactive models, with the possibility of background knowledge making up for the lack of language proficiency (Stanovich, 1980). The second approach posits the relationship between language proficiency and background knowledge in terms of defining a L2 proficiency threshold that readers need to reach in order to be able to use their background knowledge (Clapham, 1996; Ridgway, 1997; Krekeler, 2006).

In terms of orientation and methodology, studies of background knowledge have been conducted from the process and product angles. Process-oriented studies largely focused on strategy deployment, using methodologies based on think-aloud protocols and analysing qualitative data. On the other hand, product-oriented studies examined reading comprehension measured by various types of tests or recall, which were analysed from the quantitative point of view.

2.4.3 Research focus of background knowledge studies in L2 reading

Empirical research into the role of background knowledge in L2 reading has largely focused on the comparison of the relative contribution to reading comprehension of various types of background knowledge on the one hand and L2 proficiency on the other. In her review of background knowledge studies, Bernhardt (1991) divides them according to two types of background knowledge they look into: "cultural background" studies or "topic background" studies, with the subjects classified according to their ethnicity or academic background on the one hand, and proficiency level on the other. Bernhardt adds a third strand of background knowledge studies that "investigated the manner and type of background knowledge that can be given to readers in order to increase comprehension" Bernhardt (1991, p. 34). The differentiation of the studies is therefore not only based on the type of knowledge readers possess, but on the knowledge provided or activated by the researcher.

In the following sections, I survey empirical studies that examine three major sets of research foci: first, the relationship between linguistic proficiency and background knowledge with respect to the threshold above which the effect of background knowledge can be observed; second, the role of background knowledge activation; and third, the effect of topic and domain knowledge in in L2 reading.

2.4.4 Threshold studies of background knowledge

The impetus for these studies came from the conviction that readers are able to use their background knowledge and reading skills only if they reach a certain level of proficiency. A number of studies have explored whether it is possible to determine this proficiency level or threshold. Carrell (1983) found that whereas background knowledge significantly affected the comprehension of native readers, there was no such effect in the L2 group. She interpreted this result as the difference in processing by native and non-native speakers. She also found that more proficient L2 readers used their background knowledge (i.e. defined as "familiarity with the content" by Carrell) better than less proficient readers. Carrell interpreted this finding as corroborating Clarke's linguistic threshold hypothesis (1980), which claims that the transfer of reading skills from L1 to L2 can be "short-circuited" or impaired by poor L2 knowledge. The lack of L2 prevents the readers from actively using their background knowledge. According to Carrell, L2 readers tend to be more linguistically bound to the text and do not appear to make connections between the text and background knowledge. Carrell's findings had implications for the teaching of reading and language, and served as motivation for further research that attempted to specify the proficiency threshold.

In his study exploring the interaction between topic knowledge and language proficiency, Ridgway (1997) posited both a lower and an upper threshold hypothesis. He suggested that background knowledge effect is observable between the two linguistic thresholds. Below the lower threshold, the linguistic proficiency in relation to the text is too low to enable readers to use their background knowledge to interpret the text. Above the higher threshold of linguistic proficiency in relation to the text, readers are proficient enough to rely only on linguistic knowledge in order to understand the text. The findings of Ridgway's study confirmed the existence of the lower threshold, but failed to find support for the upper threshold. In addition, the results were mixed for different texts. Texts from business, architecture, and sociology were used in the experiment; the lower threshold was confirmed for the architecture text, but not for the business text, despite the fact that the texts were assessed for equivalency on three different readability tests. The test of comprehension for each text involved choosing title and headings, cloze summary, comprehension questions, a matching task, reference questions, and questions about the writer's attitude. Seeking an explanation for the results, Ridgeway looked into possible text-bound reasons such as greater difficulty of the architecture text, and the more general character of the business texts. He suggested that an educated layperson is more likely to be familiar with concepts in business than in architecture because of the influence of media and international business. In addition, he reported that the text on architecture may have been more culturally bound, having references to Western art and architectural styles. Notably, Ridgway underscored the fact that interpretation of linguistic thresholds should be related to the specific text at hand by saying:

> "Whilst background knowledge is always at work, the effect is not always detectable. It appears that a reader must be at a certain level with relation to the text for the effect to be observable." (1997, p. 161)

The observable effect of background knowledge could be attributed to readers' L2 proficiency in relation to the selected reading texts. If texts fall between the thresholds, readers use inferential processes based on topic knowledge. This enables comprehension and learning of new language items. Texts that fall within this band appear to aid readers' development most, so content-based reading syllabuses may draw on readers' background knowledge and motivation and thus facilitate learning. Ridgway makes the point that the threshold concept always refers to a particular text and an "educated L1 layman's" comprehension. This claim has two implications: first, there is no absolute threshold, but rather a measure which is relative to each particular text. Urquhart and Weir (1998) argue that thresholds thus need to be reset for each text. Second, the educated layperson reference suggests the exclusion of highly specific texts, as Ridgway limited his hypothesis to the understanding of those texts that would still be within the grasp of an educated L1 layperson and did not refer to domain-specific texts. When reading highly specific texts, educated layperson may make sense of syntax and some lexis, but cannot fully understand the text on a conceptual level (Urquhart & Weir, 1998; Clapham, 1996).

The L2 threshold was further explored and tested in a number of studies (Taillefer, 1996; Clapham, 2000; Krekeler, 2006). Clapham (1996) detected both lower and upper thresholds, implying that background knowledge can contribute to L2 reading only within a specific proficiency range. The study sample was 842 university-level EFL students divided into three language proficiency groups based on grammar test scores. The results showed that background knowledge influenced reading scores only within a specific proficiency band between scores of 60% and 80% on the grammar test. Clapham interpreted this result as suggesting two linguistic thresholds that affect the deployment of background knowledge. Students of intermediate linguistic proficiency tend to benefit from their background knowledge, students below the lower threshold (i.e. 60%) lack the language to benefit from their background knowledge, whereas students above the upper threshold (i.e. 80%) are so proficient that their proficiency can compensate for the lack of background knowledge, if the reading texts are outside of their discipline. This group of readers had similar reading scores both within and outside of the discipline. In sum, evidence from Clapham's study suggests that L2 knowledge, in particular structural knowledge operationally defined by a score on a grammar test, appears to influence the effect of background knowledge on reading comprehension. When interpreting the results, Clapham argued that the effect of background knowledge in her study may have been influenced by the insufficient text specificity.

Krekeler (2006) carried out a study that confirmed a strong effect of background knowledge, but failed to confirm the two-threshold effect. The study involved 486 international learners of German as L2 at German universities. Krekeler used two subject-specific texts that were tested for readability indices and lexical and syntactical factors. L2 language proficiency was tested with two C-tests, and background knowledge was assessed by a combination of before, after, and course measures. "Before" measures involved explaining key terms, "after" measures were readers' self-reports of topic familiarity, and "course" measures were the readers' future study fields. The ANOVA results revealed a strong effect of background knowledge on reading comprehension, as students who possessed background knowledge outperformed those that lacked it. However, with respect to thresholds, the findings were inconclusive. Students with low L2 proficiency had similar results within and outside of their future study discipline. A slight interaction between L2 proficiency and background knowledge was only detected in the test based on the business text for the students with medium level of L2 proficiency. Krekeler interpreted this finding as corroborating Clapham's results, indicating that students with grammar score between 60–80% can benefit most from their background knowledge. Nevertheless, evidence in Krekeler's study showed that most readers could use background knowledge regardless of the level of their L2 proficiency, therefore Krekeler could not support the two-threshold hypothesis. He concluded that the contributory factor to these results could be low variation in the level of both L2 proficiency and background knowledge among readers on the one hand, and a relatively low subject specificity of texts on the other. Based on the findings of his research, Krekeler expresses doubt in the existence of thresholds as clear cut-off points at which the effect of background knowledge changes significantly. He claims that if thresholds exist, they are "fuzzy" and they depend on text specificity (2006, p. 123).

2.4.5 Activation studies of background knowledge

These studies are founded on the assumption that there are two sources of background knowledge interacting in the reading process: the one assumed by the text, and the one possessed by the reader. Insufficient congruence or overlap between the background knowledge from both sources may result in comprehension problems. These problems may be compounded if the text is not explicit and it therefore requires a lot of inferred meaning. The background knowledge assumed by the text and the background knowledge possessed by the reader may never be completely congruent, as every reader brings to the text different experience and knowledge. However, if there are major mismatches between the knowledge assumed in the text and the knowledge possessed by the reader, this may impair comprehension. Activation studies raise the question as to what triggers or activates the use of appropriate background knowledge in the reading process. The underlying reasoning is that absence of relevant knowledge may not be the only problem. Readers may possess relevant background knowledge, but fail to activate and use it in the reading process. In both cases, either if they lack knowledge or fail to activate it, readers are likely to build a mental representation of the text by distorting the meaning intended by the author in order to accommodate unknown information.

Activation studies generally involve some kind of intervention on the part of the researcher, who provided various forms of activation to trigger comprehension. Activation forms range from pictorial to verbal. Carrell 1983 experimented with providing context in the form of a title or picture to accompany two highly ambiguous texts used before in L1 background knowledge studies by Bransford and Johnson: “The Washing Clothes” and “Balloon Serenade”. Hudson (1982) used pictorial and vocabulary clues. Adams (1982) used what she called “script activators”, or hints about the topic of the passage to be read. Oded and Stavans (1994) focused on how pre-reading comprehension questions shape readers’ perception of author’s view. Other activation forms used in these studies included predicting text content (Ogle, 1986), pre-reading discussion of readers’ knowledge and experiences connected to the text topic (Alvermann et al., 1985), and concept mapping (Amadieu et al., 2015).

Hudson's (1982) study explored how readers of different levels of L2 proficiency are affected by different forms of background knowledge activation. Readers were given three types of clues: pictures, a vocabulary list with definitions, and an opportunity to reread the passage. Activating background knowledge in all three ways was found to be effective, though not to the same extent across all levels of proficiency. Picture clues improved the comprehension of elementary and intermediate students, but they had no effect on the advanced group. This group had a better result when the rereading treatment was used. Hudson suggested that students of different proficiency levels vary in their abilities to form schemata from the printed input, with the advanced students being able to bring more non-visual information to the process. Nevertheless, Hudson (1982, p. 198) argued that activation of relevant background knowledge may help low-level students to mitigate the effects of their proficiency level.

The effect of background knowledge activation was also explored with respect to the language of the text (L1 or L2) and unfamiliar vocabulary. Adams (1982) used six manipulated texts containing nonsense words. Students' score of unfamiliar vocabulary was operationalized as "the number of times the target word was correctly recognized by each subject for all six passages" (1982, p. 157). Significant differences were found between the mean vocabulary scores with respect to the language variable and the presence of script activator. Higher vocabulary scores were achieved by students reading in L1, as well as those who received script activators. The author suggests that low-level L2 students may benefit more from script activators than advanced students. She concludes that the higher the proficiency level, the smaller the influence of script activators, as students are able to draw on linguistic clues from the text.

The effect of wrong or correct activators of background knowledge was tested by Oded and Stavans (1994). They carried out a study using false and correct pre-reading questions and asked readers to answer pre-reading questions and produce a summary. The authors formulated two hypotheses: first, the right schema activation will result in summaries consistent with the dominant view of the text; second wrong schema activation is going to have the reverse effect and result in students' summaries not consistent with the author's view. The study involved 177 ESL Israeli undergraduates of various disciplines divided into three groups according to the type of schema activation: 1) questions aligned with the dominant view in the text, 2) misleading questions focusing on less important information, and 3) no questions, just summary writing. The analysis of summaries and answers to questions was done with respect to whether the readers picked the author's dominant point of view. The authors hypothesized that answering pre-questions serves as instantiation of a particular schema that will generate a meaning construction consistent with questions. The questions were therefore hypothesized to be instrumental in driving the construction of meaning. The results confirmed authors' hypotheses and they concluded that an imposed schema shapes the mental representation of the text if there is interference by misleading questions, which may make readers doubtful or confused about the writer's point of view. The authors found that their research corroborated Anderson and Pearson's (1984) finding that if the reader comes across a discrepancy between his/her background knowledge and the new information in the text, they will either rule out the new information or modify the old. The distortions produced serve to accommodate new information within the framework of the stored information. The authors stress the teacher's role in helping students to activate the background knowledge necessary for reading a particular text. Comprehension questions are not seen only as a testing tool, but rather as a background knowledge activator that helps students shape their interpretation of the text.

A number of activation studies have confirmed the contributory effect of background knowledge activation to both L1 and L2 reading comprehension (Bransford & Johnson, 1972; Hudson, 1982; Oded & Stavans, 1994; Spires & Donley, 1998). The facilitative effect of activation applies only to situations when readers possess relevant and accurate background knowledge. Conversely, if readers lack relevant and accurate knowledge, the activation cannot have a beneficial effect. Alverman et al. (1985) found that if the text contained information incompatible with readers' background knowledge, this resulted in the readers' experience overriding the text information. If text information was compatible with readers' knowledge, there was no effect whether activators were used or not. Van Loon et al. (2013) studied how activation of inaccurate background knowledge affects readers' judgements. They found that inaccurate knowledge led to errors in learning new concepts. It negatively affected readers' monitoring, but it also resulted in readers' overconfidence. In their review of empirical studies of background knowledge, Cervetti and Wright (2020, p. 250) argue that the supporting effect of background knowledge activation depends on three groups of factors: 1) the level of readers' background knowledge, 2) the type of activation method, and 3) text factors, such as affirmation or refutation. They attribute inconsistencies in findings of some activation studies to insufficient alignment between readers' background knowledge and relevant textual knowledge.

Overall, research evidence generated by activation studies provides support for the facilitative effect of background knowledge activation and suggests that the failure to activate relevant background knowledge may lead to poor reading comprehension (Cain et al., 2001). For this reason, more recent activation studies set out to investigate the ways of teaching readers how to activate background knowledge. Elbro and Buch-Iversen (2013) examined how teaching readers the ways to use their background knowledge affects inferencing skill and reading comprehension scores. Their study involved 236 Norwegian sixth-grade students divided into experimental and control groups. The experimental group was trained in how to use graphic organizers (referred by authors as "graphic models") when reading 15 expository fiction and non-fiction texts, with both literal and non-literal reading comprehension questions. The control group was not given any activation instruction. The authors found a strong impact of training on gap-filling inference-making as well as on reading comprehension scores. Apart from stressing the value of background knowledge activation, the authors highlight the value of teaching readers how to actively use background knowledge for inference-making when reading expository texts.

2.4.6 Topic and domain knowledge studies

A large body of empirical studies have explored the effect of topic and domain knowledge on L2 reading comprehension and this section aims to give a survey of those that are relevant to this study. After an examination of the early studies and their characteristics, I go on to review studies covering approximately the last twenty years, which are presented in Table 2.3.

1) Early studies

Alderson and Urquhart (1985) carried out several studies testing the hypothesis that students perform better on reading comprehension tests in their own discipline than students of other disciplines, despite having the same level of language proficiency. In Study 1, the participants were four groups of students about to begin their postgraduate studies in administration and finance, engineering, maths, physics, and liberal arts. Five texts with a roughly similar Fog Index were used in the experiment: two on engineering, two on economics and finance, and one general text. Reading comprehension was assessed with a cloze test and the results showed that each group performed slightly better in their own discipline. Studies 2 and 3 were replications with the same texts and similar population. The test of reading comprehension focused on the same skills underlying reading ability: interpreting tables, interpreting anaphoric reference, and extracting main ideas. Alderson and Urquhart (1985) added short-answer questions to the cloze test to test reading comprehension. The authors' hypothesis was partly confirmed, as half of the groups scored better when reading texts in their discipline (e.g. economics), and the other half did not (e.g. engineers). The authors argue that the results show an interaction between the effect of background knowledge and language proficiency. They postulate that below a certain level of text difficulty, readers could achieve a certain reading comprehension score by means of linguistic proficiency and general knowledge. Overall, the findings of Alderson and Urquhart's (1985) three studies supported the effect of readers' discipline on their reading comprehension scores, whilst acknowledging the inconsistency of this effect and attributing it to the interaction with linguistic proficiency and the method effect.

Koh (1985) carried out a study of the effect of background knowledge on reading comprehension of Chinese ESL students of three different proficiency levels and from two different academic backgrounds: sciences and business. Four expository texts were used and were taken from business, science, history, and politics. It was assumed that first, the business text was familiar to business students but not to science students; second, the science text was familiar to science students and not to business students; and finally, the history and politics texts were classified as neutral and equally unfamiliar to all subjects. Reading comprehension was measured by cloze tests. The results showed that all groups had the highest reading comprehension score in the texts from their own discipline. Hence, Koh (1985) interpreted these results as showing a stronger effect of background knowledge than linguistic proficiency in reading comprehension. Koh (1985) acknowledged a limitation related to the operationalization of background knowledge, as after the experiment it was discovered that more than half of students had actually been science students in their secondary schooling. The author interpreted this as evidence of the fact that readers "use all resources they have in store, even in long-term memory" (1985, p. 379).

In a small-scale study, Mohammed and Swales (1984) investigated factors affecting the successful reading of technical instructions. Participants were videotaped, as the situation was a "read and do" type. Results were compared with respect to readers' linguistic proficiency and experience in scientific fields. The authors concluded that field familiarity had a stronger effect if the subjects have surpassed the threshold level of proficiency.

Bernhardt (1991) conducted a study with 63 L2 students of Spanish, using a free-association measure to assess candidates' prior knowledge, grading it on a three-point scale. The association test was administered two weeks before the experiment. Students were given five texts on the subjects they had free-associated on and were asked to recall them immediately after reading. Overall, the group scores indicated that background knowledge predicted recall, whereas the individual scores linked with background knowledge were not good predictors. Bernhardt (1991) explains that this could be due to using a three-point scale, which turned out not to be a sensitive discriminator of background knowledge. Some students who had substantial background knowledge failed to demonstrate it in recall, whereas some others who lacked it produced very good recalls. Bernhardt (1991) stresses that in similar experiments it is important to determine the particular knowledge needed, whether participants possess knowledge, and whether they manage to activate it. Bernhardt (1991) concluded that future research "must generate substantial evidence that determines whether background knowledge merely contributes to a description of L2 reading or whether it is actually a causal variable" (1991, p. 117).

Ja’far (1992) investigated the effect of science background knowledge on reading comprehension of Arabic university students of biology, chemistry, physics, maths, technology and English, all of whom are taught their disciplines in English. ELTS reading tests for General Academic, Physical Science, Life Science, and Technology were used. The tests scores showed that students performed significantly better on tests within their own major than outside of it. The author sees the results as evidence supporting the hypothesis that students’ reading performance is affected by their background knowledge. As a limitation, Ja’far notes that there was a great variation in the specificity and difficulty of individual test sections that may have affected the results. This leads to the conclusion that any selection of ESP reading material should take account of the level of difficulty and familiarity with the content.

Roller and Matambo (1992) investigated the effect of background knowledge on L2 reading comprehension of Zimbabwean bilingual readers. They replicated Bransford and Johnson’s study, using the “Washing Clothes” and “Balloon Serenade” texts. Students produced written recall immediately after reading in L1 and L2. The results showed that students performed better in L2 than L1, an effect which the authors attributed to the fact that Shona participants were more literate in L2 than in L1. The subjects also performed better on recalls of the unfamiliar “Balloon” serenade passage. This was explained as resulting from the novelty effect, tight text structure, and abstractness of nouns used in the “Washing Clothes” passage. The effect of context was detected only in the “Washing Clothes” passage. The study confirmed Carrell’s findings (1983).

One of the most extensive studies of the role of domain knowledge was conducted by Clapham (1996). The impetus behind the study was a reassessment of the question of whether it is viable to have different IELTS reading modules for students of different subject areas. The findings of this study are particularly salient for understanding the role that background knowledge and linguistic proficiency play in reading comprehension, in particular with reference to subject specificity of texts. In the main study, comprehension scores of students reading within their own subject fields were significantly higher than those reading outside of them. However, there was a difference in results between the pilot and main study, which Clapham (1996) attributed to variation in the subject specificity of texts. After running repeated measures of variance on reading subtests, she concluded that four out of ten texts were either too general or too specific. In addition, she investigated test items, but found no variation. For this reason, Clapham (1996) argued that variation in the reading passages actually contributed to the variations in specificity. In the main study, the texts at the utmost ends of the specificity continuum were removed and a clearer picture emerged, with students scoring better in the reading modules within their own field. This led Clapham (1996) to conclude that only if text specificity is controlled for and reading texts are highly specific can the effect of students' discipline be observed.

Having found that text specificity determines the background knowledge effect, Clapham (1996) set out to define the construct of text specificity, by answering the question of what makes texts subject-specific. Her study identified determining factors, but also pointed out problems. Despite surface similarity among texts and the same sources of texts, she argues that texts may still vary in specificity. Moreover, there may also be within-text variation in specificity, meaning that not all parts of the same text are equally specific. Clapham (1996) determined that text specificity largely stemmed from two factors: first, the rhetorical function of various parts of the text; and second, the amount of unexplained vocabulary, in particular "the extent to which comprehension of the text requires knowledge of subject-specific concepts which are not explained in the text." (1996, p. 191). As to the rhetorical function of a passage, descriptions of processes in research papers turned out to be subject-specific, whereas introductions to research papers did not. Further, the rhetorical function of different parts of texts appeared to indicate subject specificity better than the source of the text. Articles taken from the same source, academic journal or a book, tend to differ in specificity. Therefore, judging specificity by the source alone may not be enough. Regarding the assessment of texts according to unexplained concepts, Clapham (1996), however, reports some disagreement among her expert raters, leading her to admit that she cannot completely confirm her second claim about text specificity. Rather than from the raters, she found a confirmation of her claim about the conceptual characteristics of texts in students' assessment of the familiarity with the topic of the passage, showing that "subject area of subject specific passages was more familiar to students within the subject area than to those outside" (1996, p. 192). In her research, Clapham (1996) also compared scores from the original reading modules consisting of all texts, general and specific, and revised modules that only contained highly specific texts. She ran two multiple regression analyses and found that in the original modules, language proficiency accounted for 44% of the variance, and background knowledge raised it to 45%, indicating that test takers' reading comprehension was mostly related to their L2 knowledge. However, repeated analysis of variance run only on the highly specific texts of the revised module showed a different picture. Language ability accounted for 26% of the variance, whereas background knowledge raised it to 38%. Clapham (1996) interpreted this result as evidence showing that at higher specificity of texts, the effect of background knowledge increases. Clapham's conclusions (1996) concerning IELTS modules stressed that, as most of the tests in the IELTS battery are not subject specific despite their "surface specificity", there was no support for keeping them as modules for different academic disciplines. She argued that passages for IELTS modules should come from academic sources in a variety of disciplines, but should be general rather than subject specific, so that they are equally difficult for students from all disciplines.

2) L2 reading studies of topic and domain knowledge since 2000

This section discusses the main issues addressed in L2 reading studies of topic and domain knowledge since approximately the turn of century. Table 2.3 profiles the research focus and findings of these studies. It also gives an outline of texts, subjects, operationalization of variables, and analyses used in these studies.

a) Research focus

Overall, the surveyed studies aimed to determine the contribution of topic and domain knowledge in L2 reading in comparison to L2 proficiency as well as other variables related to the reader, text, and task. More specifically, studies examined reader-related variables including L2 vocabulary knowledge (Chou, 2011; Rydland et al., 2012), word decoding (Hwang & Duke, 2020; Rydland et al., 2012), strategies (McNeil, 2011), reading goal (Horiba & Fukaya, 2015), working memory (Shin et al., 2019), readers' affective factors such as topic interest, enjoyment, and reading motivation (Eidswick, 2010; Brantmeier, 2003; Hwang, 2019; Hwang & Duke, 2020), text factors including lexical coverage (Song & Reynolds, 2022), and finally, task factors such as multiple-text comprehension (Karimi, 2017) and question types (Burgoyne et al., 2013).

b) Participants and texts

The participants in these studies were generally ESL university students divided into groups according to their study discipline. In some studies, the aim was to compare the results between L2 and L1 readers (Rydland et al., 2012; Hwang, 2019). Texts used for L2 reading comprehension in these studies ranged from discipline-related expository texts to narrative texts. Some texts were adapted by the researchers while others were taken from the TOEFL reading battery. However, texts were seldom assessed for their specificity contrary to the suggestion made by Clapham (1996).

c) Operationalization of knowledge and reading comprehension

When discussing L2 reading studies of topic and domain knowledge, it is of prime importance to examine how they operationalized the knowledge variable because operationalization influences the results. Domain and topic knowledge was operationalized in a number of ways, most frequently as a dichotomous variable indicating subjects' discipline of study. Readers were grouped as within or outside of a particular study discipline in relation to the text. Group membership typically served as a base for the presumption of the presence or absence of topic and discipline knowledge. Despite its common use, there are two problems with this operationalization. On the one hand, it does not provide detail about readers' level of knowledge (Clapham, 1996; Horiba & Fukaya, 2015) and on the other, it may not capture all the variations in knowledge of the readers outside of the discipline, who may still be familiar with a topic and have certain discipline knowledge.

The second most common measure used for the assessment of knowledge were readers' self-reports. They provided information about readers' perception of their familiarity with the topic or discipline. They were either dichotomous (i.e. familiar vs unfamiliar) or assessed on a 5-point Likert scale. It is noteworthy that self-reports give a perception of readers' familiarity rather than an objectively measured assessment of one's knowledge. More recently, operationalization of knowledge has been based on more objective measures, such as tests of background knowledge with different question types, ranging from multiple-choice questions to short-answer questions and combinations. Some studies also used pre-teaching of background knowledge, which served as a basis for dividing readers into groups with or without knowledge (Chou, 2011).

Apart from the differences in operationalization of background knowledge, L2 studies used various operationalizations of reading comprehension. The most frequent measures are recall (Brantmeier, 2003; Horiba & Fukaya, 2015) and summary writing (Lin & Chern, 2014). Based on the conviction that there is no one best method for assessing reading (Alderson, 2000), some studies combined different types of reading assessment. Uso-Juan (2006) and Lahuerta Martinez (2013) used a combination of six different task types to assess reading comprehension, including matching words to definitions, identification of referents, text-explicit questions, true-false questions, multiple-choice questions, open-ended questions, and summary questions.

d) Results concerning the relationship between background knowledge and L2 proficiency

Overall, the surveyed research studies offer ample empirical evidence supporting the claim that topic and domain knowledge is a predictor in L2 reading comprehension (Clapham, 1996; Brantmeier, 2003; Uso-Juan, 2006; Karimi, 2017; Lahuerta Martinez, 2013). However, when background knowledge was studied in comparison with L2 proficiency, the results were mixed. Some studies found that the relative effect of background knowledge in L2 reading comprehension was lower than the effect of L2 proficiency (Clapham, 1996; Uso-Juan, 2006), while other studies found background knowledge to be a stronger predictor of reading comprehension than L2 proficiency (Lin & Chern, 2014). More specifically, in her study, Uso-Juan (2006) determined that background knowledge accounted for 21–31% of variance in comparison to 58–65% for L2 proficiency. Clapham's (1996) study discovered an effect for both L2 proficiency and background knowledge, but this effect was different in the pilot and main studies. In the pilot study, the regression model showed that L2 proficiency accounted for 44% of variance, while background knowledge raised it by 1% to 45%. However, when Clapham ran regression analyses only with texts with higher subject specificity, the contribution of background knowledge increased, so that L2 proficiency accounted for 26 % and background knowledge raised it by 12% to a total of 38% of variance. This result led her to conclude that the relative role of background knowledge in comparison to language proficiency in L2 reading crucially depends on the subject specificity of texts.

More aspects of the relationship between background knowledge and L2 proficiency were investigated in empirical studies of L2 reading that addressed the questions of whether the effect of background knowledge is the same for readers of all proficiency levels and whether readers need to reach a particular L2 proficiency threshold to be able to use and benefit from their background knowledge. In addition, these studies sought the answer to the question whether there is any compensation possible between L2 proficiency and background knowledge, or in other words, whether readers can compensate for their lower background knowledge with higher L2 proficiency or vice versa. The findings were not consistent. While some studies found no compensation effect (Lin & Chern, 2014), others detected the possibility of compensation (Clapham, 1996; Uso-Juan, 2006; Hwang & Duke, 2020). When Clapham (1996) ran a regression analysis only with high specificity texts in her study, she found a different effect of background knowledge for students at different levels of L2 proficiency. In readers with a score below 60% on a grammar test, no background knowledge effect was found, and likewise above 80%. This suggests that in Clapham's study only readers with approximately intermediate proficiency of 60–80% benefited from their background knowledge. Readers below 60% could not use their background knowledge due to language deficiency, whereas readers with L2 proficiency scores higher than 80% could use L2 proficiency to compensate for lower background knowledge. The evidence from her study led Clapham (1996) to confirm the effect of compensation as well as the lower and upper threshold.

e) Results concerning reader, text, and task factors

Although the relationship between L2 proficiency and background knowledge has been the key focus of L2 reading research, both variables have also been examined in relation to other factors. The contribution of background knowledge to L2 reading has been compared to the effect of readers' L2 lexical knowledge. In a study involving 159

Taiwanese college students, Chou (2011) compared reading comprehension results between three groups, one was given vocabulary instruction before reading, one was provided with background knowledge, and the control group had no intervention before reading. The results showed that pre-teaching vocabulary had the strongest influence on reading scores, and background knowledge had a weaker effect, just slightly stronger than no treatment in the control group. The results led Chou to conclude that effective L2 reading comprehension crucially relies on L2 lexical knowledge, therefore vocabulary building is of vital importance for L2 reading. He also noted that the background knowledge effect may be lower for longer more difficult texts. It is noteworthy that Chou (2011) drew a distinction between background knowledge and topic familiarity. He concluded that in comparison to topic familiarity, background knowledge also includes the knowledge of terminology that helps readers to make inferences in the reading process. Chou (2011) acknowledged that the subjects in his study may have only been familiar with the topic and did not possess relevant background knowledge, which resulted in a lower impact of this variable.

The role of lexical and topic knowledge as variables in L2 reading was compared by Rydland et al. (2012) and their findings contradict those of Chou (2011). Their study involved 67 Urdu and Turkish students from Norway and it compared the contribution of L1 and L2 vocabulary, topic knowledge, and word decoding in L2 reading. Vocabulary was measured with respect to its breadth in L1 and L2 as well as vocabulary depth and vocabulary cohesion function in L2. Topic knowledge, measured by a test of knowledge, was found to be the strongest single factor affecting students' comprehension; it correlated with L2 vocabulary depth, but not with decoding. Importantly, the study detected an interaction between topic knowledge and L2 vocabulary. It indicated that L2 vocabulary depth moderated the contribution of topic knowledge. In other words, readers' vocabulary knowledge influenced how students used their topic knowledge in reading comprehension. This finding importantly underlines the relationship between L2 vocabulary knowledge and topic knowledge by determining the interdependence between them. The relationship between vocabulary and topic knowledge in L2 reading was also examined by Horiba and Fukaya (2015). Their findings indicated that readers' topic familiarity facilitated conceptual processing and incidental vocabulary learning from text. They suggested that topic familiarity and language proficiency together with the reading goal affected readers' resource allocation.

Background knowledge was also examined in relation to readers' allocation of cognitive resources in the reading process. Shin et al. (2019) investigated the interaction between working memory and background knowledge in L2 reading. The study was conducted with 79 Korean EFL learners who rated their familiarity with four topics related to four reading passages adapted from TOEFL tests. The findings showed that readers with higher working memory capacity (WMC) made a better use of their background knowledge and achieved higher reading scores than readers with low working memory although both high- and low-working memory students had similar L2 proficiency levels. The study highlighted the facilitator role of working memory in the deployment of background knowledge. It revealed the ways working memory facilitates readers' use of knowledge, as high-working-memory readers were found to approach texts more strategically and were able to suppress irrelevant information. It transpires from this study that whether readers can benefit from their background knowledge depends on the constraints of their working memory. This led the authors to conclude that providing and activating readers' background knowledge on unfamiliar topics can lower their working memory load and consequently improve their reading comprehension.

Another research focus in the surveyed studies was the effect of strategies and background knowledge. McNeil (2011) conducted a study that set out to compare the individual and combined effect of background knowledge and reading strategies in L2 reading. The study involved 20 tertiary level EFL students of similar intermediate L2 proficiency; they read one text in the preliminary study and one in the main study. The combined operationalization of background knowledge and reading strategies together in the form of "self-questioning" was found to account for 56% of variance in reading comprehension scores. Individual operationalization of background knowledge in the second regression model only accounted for 0.4% of variance in reading comprehension scores. This led to the conclusion that the contribution of background knowledge to reading comprehension was weak in this study. The author attributed it to the intermediate L2 proficiency of readers, in combination with difficult texts, which led to a similar effect as in studies with low proficiency readers, in which background knowledge was not found to support reading comprehension.

Apart from readers' cognitive factors, the role of background knowledge was also investigated in relation to readers' affective factors in the L2 reading process. Eidswick (2010) examined the effect of readers' interest and prior knowledge in L2 reading comprehension. His study included 23 Japanese university students who read expository texts on eleven topics. Knowledge and interest were measured by readers' self-reports and readers were divided into groups according to their levels of knowledge and interest. The ANOVA results revealed significant differences between the "high interest-high knowledge" group and both groups with low knowledge, irrespective of their interest level. The evidence led the author to conclude that prior knowledge played a different and stronger role in reading comprehension than interest. Nevertheless, Eidswick (2010) argued that the joint contribution of background knowledge and interest represents a most powerful combination in reading comprehension. Similar results were obtained by Hwang (2019) in a study comparing readers' affective factors and background knowledge. Specifically, the study investigated the role of science domain knowledge and reading motivation in L1 and L2 reading comprehension of informational and narrative texts, by involving L2 and monolingual fourth-grade students in five countries. The ANOVA results showed that science domain knowledge was a stronger predictor of reading comprehension than reading motivation both in L1 and L2. The role of science domain knowledge was found to be equally or even more important in L2 comprehension than L1 comprehension.

In addition to a range of cognitive and affective reader factors, L2 reading studies examined various text and task factors in relation to background knowledge. Among the text factors, studies focused on the impact of genre, in particular expository texts (Song & Reynolds, 2022; Lahuerta Martinez, 2013; Eidswick, 2010; Uso-Juan, 2006) and narrative texts (Hwang, 2019; Horiba & Fukaya, 2015), their rhetorical functions and relationships within texts involving cohesion and coherence, and text difficulty (Droop & Verhoeven, 1998; Kobayashi, 2002). Task factors mostly referred to question types used in reading comprehension, especially text-explicit, text-implicit, and script-implicit questions. Although a large number of L2 reading studies involving the background knowledge variable used multiple texts with reading comprehension, each text was assessed separately. In contrast, Karimi (2017) conducted a study that compared single-text comprehension and multiple-text comprehension. Multiple-text comprehension went beyond reading a number of individual texts because it required the readers to connect information across texts. The main focus of Karimi's (2017) study was to determine the role of L2 proficiency and background knowledge in both single and multiple mode of text comprehension. The regression results revealed that both L2 proficiency and background knowledge contributed to multiple-text comprehension; however, L2 proficiency was a stronger predictor in both the single and multiple modes of text comprehension. Background knowledge accounted for more variance in multiple rather than single text comprehension, which was explained to be a result of the fact that single-text comprehension actually mediated multiple text comprehension.

Table 2.3 Empirical studies of topic and domain knowledge in L2 reading

Study	Text	Subjects	Research focus BK operationalization RC operationalization Analyses	Results
Clapham, 1996	10 passages, for 3 discipline combinations: business and social sciences, life and medical sciences, physical science and technology	842 university EFL students, divided into three language proficiency (LP) groups according to grammar test scores.	To determine the effects of discipline-specific BK on RC. BK dichotomous – within or outside of discipline & reading habits & subject familiarity Multiple regression and ANOVA used.	Readers had better RC within their discipline if the texts were highly specific. In the main study, LP accounted for 26% of variance, BK for 12%, hence LP was found to be a stronger predictor of L2 RC than BK. Confirmed lower and upper L2 proficiency thresholds of 60–80% grammar test score. Readers with scores below 60 or above 80% did not benefit from their BK in their RC.
Brantmeier, 2003	2 passages from short stories	85 university students of Spanish	To investigate the effects of readers' gender, topic familiarity, enjoyment, and interest on L2 reading. Topic familiarity assessed by self-reports on a 5-point Likert scale. RC operationalized as free written recall. ANOVA used.	In addition to language factors, readers' reading comprehension was affected by passage content and topic familiarity.
Uso-Juan, 2006	6 expository texts: 2 from psychology, 2 tourism, 2 engineering	380 Spanish EFL learners, divided into 3 groups: psychology, tourism, engineering	To determine the contribution of discipline-related knowledge and language proficiency in EAP reading. Discipline knowledge assessed by a 10-item MCQ test. RC assessed by 6 different tasks. Regression used.	Both LP and discipline-related knowledge were predictors of RC, LP accounted for 58–65% of variance, and discipline knowledge for 21–31%. A compensatory effect was detected: readers could compensate for the lack of BK with higher LP.

Study	Text	Subjects	Research focus BK operationalization RC operationalization Analyses	Results
Eidswick, 2010	Expository texts on 11 topics	23 Japanese university students of an English program	To examine interest, prior knowledge in relation to comprehension. Knowledge and interest rated by self-reports. RC assessed with MCQ test. ANOVA used.	Significant differences in RC found between "high interest - high knowledge group" and "low knowledge – low & high interest groups". BK had a stronger influence on RC than interest.
McNeil, 2011	2 adapted texts	20 tertiary-level EFL students	To determine the individual and combined effects of BK and reading strategies on L2 RC. BK assessed with rating 12 main idea statements in a questionnaire by using a 4-point scale to rate familiarity. Regression analyses were used.	BK and RC strategies were jointly, operationalized as self-questioning. Self-questioning accounted for a large part of variance in RC scores, whereas BK was not a strong predictor of RC. RC strategies accounted for substantially larger share (56% vs 0.4%) of variance.
Chou, 2011	3 reading passages from a TOEFL test	159 EFL Taiwanese college students divided into 3 groups.	To examine the effects of vocabulary knowledge and BK in RC. Before RC one group was given BK information on text topics, one group a vocabulary list and control group no treatment. BK assumed after pre-teaching. ANOVA was used.	The "vocabulary group" scored significantly higher than control group and "BK group". The BK group scored slightly higher than the control group. Vocabulary pre-teaching was found to have stronger influence on RC than BK.
Rydland et al., 2012	3 texts on global warming constructed by researchers	67 Urdu and Turkish L2 fifth-grade students in Norway	To examine the contribution of word decoding, L1 and L2 vocabulary and prior knowledge in L2RC. BK test with MCQ and SAQ. RC test with MCQ	Topic knowledge was the strongest predictor of RC. There was interaction between topic knowledge and L2 vocabulary. L2 vocabulary depth moderated the contribution of topic knowledge.

Study	Text	Subjects	Research focus BK operationalization RC operationalization Analyses	Results
			and SAQ. Multiple regression was used.	Topic knowledge was not correlated with word decoding but with L2 vocabulary.
Lahuerta Martinez, 2013	1 narrative text from ESOL examinations	129 intermediate university Spanish L2 students	To analyse the effect of perceived interest and prior knowledge on L2 RC. Topic familiarity self-reported on a 5-point scale. RC assessed by recall, cloze and MCQ. Regression used.	Prior knowledge had a positive effect on RC irrespective of type of RC test. RC is improved if texts are related to readers' interests.
Lin & Chern, 2014	TOEFL RC test	71 Taiwanese EFL learners	To investigate how L2 knowledge and BK influence RC. Discipline-related BK test - SAQ. RC- summary writing in L1. ANOVA was used.	Both L2 proficiency and BK affected RC. BK was a stronger predictor of RC than L2P. There was no compensation between L2P and BK.
Horiba & Fukaya, 2015	2 narrative passages on a healthcare and nursing	145 EFL college students divided into 2 groups within and outside of discipline of nursing	To examine the effects of reading goal, topic familiarity, and L2 proficiency in RC and learning. BK – dichotomous: high and low topic familiarity according to readers' discipline. RC – recall in L1 and L2 ANOVA was used.	L2 proficiency affected recall and incidental vocabulary learning. Topic familiarity facilitated conceptual processing, learning of the text-content and incidental vocabulary learning. Reading goal was found to affect resource allocation in text processing.
Karimi, 2017	4 passages with medical topics from a textbook, semi-technical book, and	207 Iranian undergraduates studying midwifery and paramedical emergencies.	To explore contributions of LP and BK to RC across L2 multiple-text and single-text modes of reading. Used a 20-point RC test that required the readers to connect	LP and BK were found to contribute to multiple-text comprehension. The contributions varied across single-text and multiple-text modes of reading. LP was found to be a stronger predictor in both single- and multiple-text reading. However,

Study	Text	Subjects	**Research focus** **BK operationalization** **RC operationalization** **Analyses**	Results
	popular science.		information across four texts. BK assessed with a 30-item MCQ test. Structural equation modelling (regression) was used.	BK accounted for more variance in multiple- rather than single-text comprehension (39% vs 26%). There was a mediation role of single-text comprehension in multiple-text comprehension.
Hwang, 2019	Informational and narrative texts	Two groups: fourth–grade L2 students and monolingual students in five countries	To investigate the role of science domain knowledge and reading motivation in L1 and L2 RC of informational and narrative texts. Science knowledge test. Multi-group multi-level regression was used.	Science domain knowledge was the strongest predictor of RC for both L2 and monolingual readers in all five countries. Science-domain knowledge was equally important in L2 and L1 RC. Contrary to other research it seemed even more important in L2.
Shin et al., 2019	4 passages adapted from TOEFL practice tests on four different topics	79 adult Korean EFL learners	To explore whether and how working memory and background knowledge combine to facilitate L2 RC when readers have or do not have background knowledge. BK: self-rated familiarity with four topics on a 5-point scale.	L2 readers with higher working memory capacity (WMC) benefited more from the provided BK than low WMC readers and also achieved better RC results. L2 proficiency results were similar between readers with high and low WMC, which underscored the role played by WM. The results support the rich-get-richer model, positing that the effect of background knowledge can be strengthened by high WMC. High WM students appeared to approach the RC task more strategically by suppressing irrelevant information. Readers can benefit from BK if they have good WMC, otherwise WMC can hinder the use of BK.
Hwang & Duke, 2020	Texts and tests from the national	12,101 third-grade ESL and	To examine the contribution of science domain knowledge,	RC was significantly associated with science domain knowledge, motivation for reading, decoding

Study	Text	Subjects	Research focus BK operationalization RC operationalization Analyses	Results
	Assessment of Educational Progress	monolingual readers	reading motivation and decoding skills in RC. Science domain knowledge test with SAQs. Regression was used.	skills and early development of decoding skills. As the coefficient of science domain knowledge was significantly higher for ELs, this can imply a compensation effect, with knowledge compensating for LP.
Song & Reynolds, 2022	2 expository texts: 1 on familiar topic, 1 unfamiliar topic	81 Chinese EFL undergraduates	To examine the effect of topic knowledge and lexical coverage on RC of L2 expository texts with two different lexical coverages: 95 and 98%. Topic familiarity assessed with MCQ test. Two-way ANOVA was used.	The effect of topic knowledge on RC was significant and higher than the effect of lexical knowledge. Topic knowledge improved RC irrespective of the lexical coverage. It was also found that topic knowledge helped readers to suppress irrelevant information.

2.5 A summary of the role of background knowledge in empirical studies of L2 reading

Both theory and empirical research offer ample evidence supporting the contributory role that background knowledge plays in reading comprehension. Theoretical justification for the inclusion of background knowledge in reading process research comes from three sources: interactive and componential reading models, schema theory, and most recently, construction-integration models. Interactive reading models (Rumelhart, 1977; Just & Carpenter, 1980; Rayner et al., 2011), posited the interaction of low-level and higher-level processes in the reading process, including background knowledge. They also argued for the possibility of accommodating for L2 deficiencies by higher level processing, thus suggesting compensation. The second source of theoretical support for the inclusion of background knowledge is componential models of reading (Coady, 1979; Bernhardt, 1991), which list background knowledge as a component underlying reading together with conceptual abilities and process strategies in the Coady model and phonographic features, word recognition, syntax and perception in the Bernhardt model. Despite its theoretical limitations, schema theory (Rumelhart, 1975; Schank, 1975; Schank & Abelson, 1977), in particular, motivated a plethora of studies that included background knowledge. A more recent theoretical justification for the inclusion of background knowledge in reading studies comes from the Construction-Integration models (Kintsch, 1998; Perfetti, 1999). Among the three phases of processing they posit, the third phase referred to as the situation model, involves readers' integration of the text information with readers' background knowledge. In this process, prior knowledge supports comprehension, while comprehension contributes to building new knowledge (Cervetti & Wright, 2020).

As shown in previous sections, there is a large body of empirical research that examined and tested the role of background knowledge in comprehension both in L1 and L2 reading. Overall, empirical studies converge on the finding that background knowledge plays a facilitative role in reading comprehension. Effective comprehension therefore relies on the integration of text information with the reader's prior knowledge. A large number of studies have provided empirical evidence that readers with better topic or domain knowledge read texts more efficiently and with better comprehension. More specifically, the findings have shown that readers with more topic knowledge produced better recalls of text information as well as better answers to various types of questions, ranging from text-explicit questions to text-implicit and script-implicit questions. A facilitative effect of knowledge was also identified for readers with high domain knowledge, despite more variation in results if compared to the effect of topic knowledge (Cervetti & Wright, 2020).

It transpires from theoretical and empirical studies that background knowledge is at least a two-faceted variable that refers both to the reader and the text. The reading process involves the interaction between the knowledge the reader possesses and the knowledge the text assumes the reader to have. Alderson (2000, p. 61) argues that "what readers know will affect what they will understand when reading" and continues that "text content will affect how readers process the text." If there is a mismatch between background knowledge assumed in the text and the knowledge possessed by the reader, the reader will try to accommodate new information by distorting it, or ruling it out (Steffensen et al., 1979). It is therefore crucial to see knowledge as pertaining to three factors: the text, the reader and the task. Text-bound knowledge or the knowledge that the author assumes the reader to have can be mediated by other text factors such as the context and transparency (Carrell, 1983). Texts may provide different amounts of clues and they may range from opaque to transparent. If the text is not transparent or lacks context, the role of readers' background knowledge may be even more important. Insufficient congruence or mismatches of readers' and text knowledge may result in comprehension problems.

In L2 reading, readers' L2 proficiency is the most important mediating factor affecting the use of background knowledge. Threshold studies aimed to determine the threshold level of L2 proficiency that allows background knowledge to be used by the reader. A lower threshold was confirmed, implying that readers need to reach a particular level of L2 proficiency in order to be able to use their background knowledge (Clapham, 1996; Ridgway, 1997). The idea behind the upper threshold was that above a particular level of L2 proficiency background knowledge no longer affects reading comprehension. While Clapham found support for the upper threshold, Ridgway did not, so this question remains open (Clapham, 1996; Ridgway, 1997). In fact, Clapham found that subjects scoring between 60 and 80% on a grammar test benefited from their background knowledge, whereas students outside of this range did not benefit. Finally, when discussing thresholds, it is of prime importance to note that the L2 threshold is not an absolute level across all studies, but is a concept that needs to be determined in relation to the texts and readers in each study (Alderson, 2000, p. 39).

Another aspect concerning the relationship between L2 proficiency and background knowledge is the possibility of compensation of one variable for another as articulated in the compensatory interactive models (Stanovich, 1980). These models hypothesized that readers can compensate for lower L2 proficiency with higher background knowledge and vice versa. Only a few empirical studies have tested this hypothesis (Koh, 1985; Chen & Donin, 1997; Uso-Juan, 2006) and found some support for the compensation hypothesis, with the most compelling results in Uso-Juan's study (2006). Her findings led her to conclude that the lack of discipline knowledge can be compensated for by both intermediate and advanced level of L2. In addition, a low level of L2 proficiency can be compensated for by discipline knowledge. Finally, to have a contributory effect in the reading process, background knowledge needs to be activated and this activation was found to be particularly relevant for low proficiency students (Hudson, 1982; Adams, 1982). At higher levels of proficiency, when readers draw upon linguistic clues more successfully, activation of background knowledge no longer had the same effect. Hudson concluded that pre-reading schema activation techniques are beneficial to beginner and intermediate students in terms of enabling them to use reading strategies (Hudson, 1982).

L2 proficiency or knowledge is generally treated as a multicomponent construct and L2 reading research examines to what extent different components account for reading comprehension. In their review of multicomponent reading studies, Jeon & Yamashita (2020) argued that L2 reading comprehension is strongly associated with L2 knowledge and that this association is stronger than the association with cognitive or metacognitive variables. They note that L2 knowledge is typically operationalized as either sub-lexical, lexical or grammar knowledge. Grammar knowledge refers to the linguistic knowledge involved in the morpho-syntactic and syntactic processing of written text. It includes explicit and implicit knowledge involved in recognizing parts of speech, inflections, phrases, syntactic role of clauses, and sentence word order. Jeon and Yamashita (2014, 2020) systematically analysed the correlation coefficients and summarized the evidence for the association between L2 reading comprehension on the one hand and eleven correlates on the other, including: decoding, phonological awareness, vocabulary knowledge, grammar knowledge, L1 reading comprehension, L2 listening comprehension, working memory, orthographic knowledge, morphological knowledge, metacognition, and L2 oral reading fluency. L2 grammar knowledge and L2 vocabulary knowledge were found to be the strongest correlates of L2 reading comprehension ($r = .790$ for grammar and .724 for vocabulary) in both reviews of reading studies. This led Jeon and Yamashita (2020) to deduce that L2 grammar knowledge and L2 vocabulary knowledge are “two core L2 knowledge variables at the heart of abilities required for L2 reading comprehension.” (p. 60). They concluded that “… the most important, sustained finding is the strong relationship of L2 linguistic knowledge (especially grammar and vocabulary) and L2 reading comprehension.” (p. 72).

2.5.1 Inconsistencies in empirical studies of L2 reading

Several methodological issues or inconsistencies have affected empirical studies of L2 reading (Cervetti & Wright, 2020). These issues may have contributed to the inconsistent findings and consequent difficulty when comparing results. The inconsistencies mainly refer to two groups of factors: a) operationalization of background knowledge and reading comprehension and b) textual factors, especially text specificity.

a) Operationalization of background knowledge

Background knowledge in empirical studies of L2 reading has been operationalized in a number of ways:

1. Readers' study discipline (Alderson & Urquhart, 1983; Koh, 1985; Peretz & Shoham, 1990; Ja'far, 1992; Clapham, 1996; Horiba & Fukaya, 2015);
2. Self-reported familiarity with the topic or content (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Lin, 2002; Brantmeier, 2003; Salmani-Nodoushan, 2003; Pulido, 2007; Eidswick, 2010; McNeil, 2011; Lahuerta Martinez, 2013; Shin et al., 2019);
3. Free introspection (Bernhardt, 1991);
4. Oral interview (Chen & Donin, 1997);
5. Pre-teaching of background knowledge (Chou, 2011);
6. A combination of different measures: current, past or future study field, topic familiarity report, reading habits, and reading interest (Bügel & Buunk, 1996; Krekeler, 2008; Chigayeva, 2000);
7. Score on a test of background knowledge (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020; Song & Reynolds, 2022).

Some aspects of these operationalizations have been questioned and the critique referred to the lack of detail. The operationalization of knowledge as the readers' current, past or future study discipline is based on the assumption that readers within the discipline possess background knowledge, while readers outside of the discipline do not. This assumption may not be entirely true, as people outside of discipline can also possess some discipline knowledge. In addition, being within or outside of a discipline is a dichotomous operationalization that does not render any detail regarding the level of knowledge a reader possesses. For these reasons, the operationalization of knowledge through assumptions based on one's discipline of study has been criticized as inadequate (Bernhardt, 1991; Clapham, 1996, p. 169). Another widely used operationalization of background knowledge used was readers' self-reports of their topic or domain knowledge in which readers assess their familiarity with the topic before or after reading either dichotomously or by using a 5-point Likert scale. The main concern here is the fact that self-reporting the familiarity or knowledge is a subjective measure and may not be reliable.

Free association as operationalization of readers' knowledge taps into the totality of one's experience and knowledge, but the generated data may not be relevant or related to the text topic. While it is possible to take free association as an indication of one's knowledge, it is worth pointing out that free associating on a topic may produce results that are unrelated to the text and thus not a measure of the textually relevant background knowledge.

More recent studies have used tests of knowledge to operationalize knowledge related to reading texts, thereby assuring a more objective measure of one's knowledge (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020; Song & Reynolds, 2022). Despite some possibility of priming the results (Spiridiakis & Wenger, 1991), testing knowledge appears the most reliable and accurate measure of topic or discipline knowledge related to a particular text.

b) Operationalization of reading comprehension

There is some variation in the way reading comprehension is operationalized in empirical studies. The assessment measures range from recall or summaries, tests with different item types: short answer questions, multiple choice questions or cloze. Any of these options can affect the results. For example, recall relies on readers' memory and may therefore not only assess reading comprehension. Cervetti and Wright (2020, p. 243) argue that recall as an assessment of comprehension "reveals little about the degree of integration of textual information with the existing knowledge". In addition, reading comprehension questions can vary as to what they require the reader to do in relation to text. Cervetti and Wright (2020) distinguish between three types of questions: a) text-explicit questions, b) text-implicit questions, and c) script-implicit questions. While the text-explicit questions require the reader to find the answers in the text, text-implicit questions require the reader to make inferences based on text information and make connections within a text to fill the gaps in the text with the information that the author does not state explicitly. Script-implicit questions require the reader to use background knowledge outside the text. Question types and test formats can affect the comprehension results.

c) Text specificity

Text specificity was found to influence whether the background knowledge effect can be detected (Clapham, 1996). Reading comprehension of low-level domain-specific texts is mainly influenced by the readers' linguistic ability. On the other hand, reading comprehension of discipline texts with high specificity may not only be affected by the reader's linguistic ability. Clapham acknowledged that the texts used in her own and in a number of other studies may have been general rather than subject specific, which affected the results and particularly, the contributory role of background knowledge. This could have been further exacerbated by the fact that determining the level of text specificity may be rather subjective, despite involving subject experts. In her study, Clapham found that the rhetorical structure of texts and the amount of unexplained concepts contributed considerably to text specificity.

2.5.2 Gaps and open questions in L2 reading research related to background knowledge

Although the role of background knowledge in reading comprehension has been acknowledged by a large body of theoretical and empirical studies in both in L1 and L2, the results have been mixed as regards the relationship between L2 proficiency and background knowledge, in particular the existence of proficiency thresholds and the possibility of compensation. This led Koda (2005) to conclude her review of reading research by saying that: “Additional explorations of background knowledge, particularly the interplay with L2 proficiency during comprehension deserve further attention.” (2005, p. 152).

More recent reviews of the literature have identified several inconsistencies in L2 reading studies that may have contributed to inconclusive or mixed results (Cervetti & Wright, 2020). Inconsistencies related especially to the operationalization of the main variables (i.e. background knowledge and reading comprehension) and they were identified also by some authors of empirical studies who called for a more objective measures of knowledge in the form of tests (Uso-Juan, 2006; Shin et al., 2019; Karimi, 2017). Apart from operationalizations of variables, a major open issue identified in L2 reading studies was inadequate subject specificity of texts. The issue was raised by Clapham (1996) and she concluded that text specificity was a major factor and condition for detecting the effect of background knowledge.

Finally, the reviewers of L2 reading have noted a lack of mixed studies combining quantitative and qualitative methodologies (Grabe, 2009; Koda, 2005). Mixed studies could address not only the question of the strength of the background knowledge effect in L2 reading, but also relate it to the ways background knowledge affects the reading process. Grabe (2009, p. 76) stressed that background knowledge is “an essential component of reading and central to the construction of an appropriate situation model of text interpretation”. Nevertheless, he warned that background knowledge “is not a concept that is well-specified to this point in reading research”, arguing that it is not “transparent how background knowledge is called up and used in reading comprehension” (2009, p. 74). This observation seems to reflect Clapham’s view (1996, p. 203) that further research into background knowledge should also be qualitative and should explore how readers use their background knowledge when reading subject-specific texts.

The literature review presented in this chapter and the identified open issues informed the research questions and design of this study. Two sets of data are combined: testing scores obtained in the quantitative study and think-aloud data from the qualitative study. A review of the think-aloud methodology is presented in Chapter 3 and the study design in Chapter 4.

2.6 Research questions

This study sets out to explain how much and in what ways background knowledge affects reading comprehension, especially how it interacts with grammar knowledge. I compare groups with high and low levels of background knowledge and high and low L2 grammar knowledge with respect to their reading comprehension. Based on the interaction of background knowledge and L2 grammar knowledge, I seek to determine the existence of L2 threshold for the use of background knowledge as well as the possibility of compensation between these two variables. I use two measures of knowledge: tested Background Knowledge (BK) and Self-reported Familiarity (SRF). Background Knowledge is compared to Self-reported Familiarity as regards the impact on reading comprehension. Grammar Knowledge (GK) is used as the indicator of L2 knowledge, which is the approach taken in a number of L2 reading studies (Jeon and Yamashita, 2020). As vocabulary is a component both of L2 proficiency and background knowledge, the main aim of using Grammar knowledge is to reduce the effect of the vocabulary component and avoid confounding variables. Finally, I combine the results of quantitative and qualitative studies and use them to answer not only the question about the strength of background knowledge impact on reading comprehension, but also about the ways background knowledge is deployed in reading comprehension of subject-specific texts.

The study addresses the following research questions:

RQ 1: Do Background Knowledge, Grammar Knowledge and Self-reported Familiarity affect Reading Comprehension of subject-specific texts?

RQ 2: To what extent do Background Knowledge, Grammar Knowledge and Self-reported Familiarity explain individual differences in Reading Comprehension of subject-specific texts?

RQ 3: Does the contribution of Background Knowledge to Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

RQ 4: To what extent does the contribution of Background Knowledge to Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

> Hypothesis: Students can use their Background Knowledge only at a certain level of Grammar Knowledge

RQ 5: Can students with better Background Knowledge compensate for lower Grammar Knowledge in Reading Comprehension of subject-specific texts?

RQ 6: Can students with better Grammar Knowledge compensate for lower Background Knowledge in Reading Comprehension of subject-specific texts?

> Hypothesis: Better Grammar Knowledge allows students to compensate for lower Background Knowledge, and vice versa.

2.7 Chapter summary and conclusions

This chapter presented a survey of theoretical and empirical studies of L2 reading with regard to background knowledge. The theoretical underpinning has been given for the background knowledge as a variable in reading comprehension by reviewing reading models from the componential, linear, and interactive to construction-integration types. The review of empirical studies began by discussing the definition and operationalization of background knowledge. It continued by looking at the research foci of these studies, especially the relationship between background knowledge and L2 proficiency. Four groups of studies were reviewed: threshold, activation, topic knowledge, and domain knowledge studies. The chapter went on to provide an overall picture of findings and limitations in the reviewed studies. Finally, the research questions were formulated based on gaps identified in the literature review.

CHAPTER 3
THINK-ALOUD METHODOLOGY AND READING COMPREHENSION

3.1 Chapter aim and overview

This chapter examines the methodological issues that guided the study. It first provides an overview of the product and process approaches to researching reading comprehension, followed by a review of the theory that underpins the think-aloud (TA) methodology and a discussion of its strengths and limitations. After that, the chapter provides a survey of empirical studies using think-aloud methodology in L1 and L2 reading. The chapter concludes by proposing a methodological framework for this study and formulating research questions 7 and 8.

3.2 Product and process approaches to reading comprehension

Reading studies are designed from the starting point that reading is an invisible cognitive process, which can only be approached in an indirect way by observing the way readers read or examining what readers have gained by reading. In other words, reading studies focus either on the process of reading or the product of reading.

Outcomes or products of reading are investigated by quantitative studies that tend to operationally define reading comprehension as a score on a reading comprehension test. The reading comprehension tests comprise of different types of questions, ranging from short-answer and multiple-choice questions to gap-fill or cloze. A frequent measure is also verbal or written recall, performed in subjects' L1 or L2. The measurement of reading comprehension outcomes may be affected by the methods used. Both the cloze procedure and multiple-choice questions have drawn criticism as to their construct validity. Cloze may be a measure of clause comprehension rather than global understanding of a text. Multiple-choice questions appear to measure the skilful solution of a task rather than readers' comprehension. Recall may not only measure readers' comprehension of a text but also their memory. Even though there is a relationship between memory and comprehension, and the two processes are intertwined, they are not interchangeable. It appears that by understanding a text, one may remember it better, but a full understanding of a text may not necessarily result in perfect recall. Therefore, it appears that no single measure of reading comprehension should be used as a stand-alone indicator of a reader's comprehension, but it should rather be combined with other measures that are both product- and process-oriented.

Process approaches to reading investigate observable features of reading such as readers' eye-fixations or pauses, assuming that they indicate different aspects of the reading process, which motivated the use of eye-tracking in reading research. (Brunfaut & McCray, 2015; McCray & Brunfaut, 2018). There are, however, limitations to what one can infer about reading processes on the basis of observable features of silent reading. To deal with unobservable elements of the reading process, introspective and retrospective methods have been developed and refined (Ericsson & Simon, 1984; Pressley & Afflerbach, 1995; Israel, 2015; Bowles, 2010). These relate especially to higher-level reading processes that have not yet been automatized.

3.3 Introspective and retrospective methods

Introspective methods focus on reading processing in real time, whereas retrospective methods rely on readers' reports after reading. Introspective methods are said to offer a window into internal cognitive processes, so that one can observe "stream of consciousness disclosure of thought processes while information is being attended to" (Cohen, 1987, p. 84). Doing an introspective task, also termed a think-aloud task, requires the reader to verbalize the thoughts going through their mind when reading a text. Readers' verbal reports accompanying reading are usually tape-recorded, transcribed, and then analysed. Retrospective methods may shed additional light on the reading process. They involve readers' comments on their reading process after they have read a text. In particular, after completing the task readers report on the types of problems they encountered, how they resolved them, and the kind of knowledge they drew on.

3.4 Theory underpinning think-aloud methodology

The think-aloud methodology has been guided by theoretical and empirical studies. Here I review those theoretical studies that have mainly addressed the following areas of think-aloud methodology: a) levels of verbalization and types of verbal protocols, b) types of reported information, c) maintaining concurrency, d) training and instruction of think-aloud participants, and e) validity and reliability.

3.4.1 Three levels of verbalization and types of verbal protocols

The Ericsson and Simon (1984) theoretical model of verbal protocols is based on the conceptualisation of three factors: short-term memory, long-term memory, and metacognition. Long-term memory stores a huge capacity of different types of knowledge organized to different degrees, including declarative, procedural, and conditional knowledge. This is knowledge about things and knowledge of how and when to do things, all retrievable when needed. On the other hand, short-term memory is conceptualised as the concurrent contents of one's consciousness. It has a limited capacity, but can be accessed quickly and individuals can report about it. Ericsson and Simon (1984) therefore contend that people are able to make valid reports of their short-term memory provided that reporting takes place within the scope of short-term memory, which covers two to ten seconds after the event. They also argue that only the information in short-term memory shows internal processes. Therefore, if verbalization is to reveal these processes, it needs to take place within two to ten seconds after reading. Any report beyond this time is not a reflection of a person's current processing, but rather a retrospection on the process that reflects one's long-term memory and is likely to be influenced by other factors. Thus, it may not be accurate and it may contain explanations or subjects' reflections which were not a part of the processing itself. Therefore, Ericsson and Simon's (1984) major guideline to maintaining the validity of verbal reports concerns the time limits that enable readers to report processes still held in short-term memory. They also contend that even though reporting the contents of one's short-term memory may slow down processing, it does not affect the validity of the verbal protocols. Similarly, Olson et al. (1984) find immediacy of reporting to be crucial to the validity of subjects' reports, but they also see it as a major constraint.

With regard to the intervening processes between the reading process and verbalization, Ericsson and Simon (1984) distinguish three levels of verbalization. The first level is direct verbalization, such as naming things, which does not involve any intervening processes. The second and third levels are "encoded" verbalizations that allow for other processes to take place before the verbalization. The second level involves descriptions of one's thoughts, and the third level involves readers' explanations of how they processed written information. Ericsson and Simon (1984) argue that verbalization at the third level no longer reflects one's real cognitive process. Therefore, the validity of think-aloud data can be maintained when involving only the first- and second-level verbalizations.

In sum, Ericsson and Simon (1993) find the data obtained from verbal protocol analysis to be a particularly good source for studying the cognitive processes and strategies involved in reading comprehension. They make a distinction between the following:

a) Concurrent or think-aloud reports,
b) Stimulated recall reports,
c) Retrospective verbal reports.

Concurrent or think-aloud protocols reveal the participant's verbalization taking place simultaneously as they are processing textual information, whereas stimulated recall protocols refer to reporting on the processes immediately after completing the task when the processes are still in participant's short-term memory. Because retrospective verbal reports involve readers' reporting with a time lapse after the task completion, this implies that retrospective reports rely on the reader's memory, and this may affect the result (Bowles, 2010). For this reason, concurrent think-aloud reports and stimulated recall protocols have been found to have superior validity when compared to the retrospective ones.

3.4.2 Type of reported information

Ericsson and Simon (1984) translated the distinctions among the three levels of verbalizations into instructions for conducting verbal reports as regards the type of reported information, in particular metacognition. The participants should be told to express only what they are thinking about when doing the reading task without justifying their actions, explaining reasons, or interpreting their processing. Participants should be asked to verbalize thoughts with encouragement such as: "Try to think aloud as thoughts pass through your head." Participants might assume that they need to defend and explain the way they are doing things, so they should be alerted not to do so. Basically, if participants tailor their self-reports to the researcher's expectations, this affects the validity of reporting the process. Subjects should also not try to make their reports more coherent. Krutetskii (1976, quoted in Ericsson & Simon, 1984) offers a simple instruction to the reader: "Do not try to explain anything to anyone else. Pretend there is no one here but yourself. Do not tell about the solution but solve it." It is noteworthy that in making sense of the reports, it is the researcher who explains or interprets the processing — not the participants.

3.4.3 Maintaining concurrency

Ericsson and Simon (1984) stress that it is only possible to report processes that are not automatized, because they are slower and sequential. Similarly, Olson et al. (1984) indicate that the think-aloud method lends itself best to investigating higher-level processes such as inferencing, predicting, and schema elaborating, because these processes are not automatized and are consciously accessible. On the other hand, they find that lower-level processes — such as letter and word recognition — are fully automatized in skilled readers, and they are consciously inaccessible. For this reason, techniques to slow down reading speed are suggested as they would allow for concurrent verbalization. Texts may be presented sentence by sentence, or readers can be asked to stop and think aloud at certain points marked in the text.

Pressley and Afflerbach (1995) find that many studies appeared to have used retrospective rather than introspective reports. However, this is often difficult to figure out because studies often lack sufficient detail to make it clear what sort of report was actually used. This motivated Pressley and Afflerbach (1995) to readdress the fundamental question: what difference does it make if one uses concurrent or retrospective reports? They found that there had been little systematic comparative research of concurrent and retrospective methods; moreover, studies using verbal reports generally lack specificity concerning this aspect of methodology and therefore cannot be compared. Pressley and Afflerbach (1995) speculate that many reports may have been carried out very close to actual reading, but they may not have reported the contents of subjects' short-term memory. They believe that reports delayed more than ten seconds, as suggested by Ericsson and Simon (1984), may contain traces of original processing and could largely be seen as the reader's reconstructions of the reading process; the major difference between concurrent and delayed reports being the number of processes following the reading task.

Continuing the discussion of concurrency of verbal reports, Pressley and Afflerbach (1995) also argue that there may not be such thing as a 100% concurrent report, because any verbal report can only be given after the completion of at least a small section of reading. This makes them conclude that there may not be a great difference between a "concurrent" and "briefly retrospective" report. For them the critical question is how long reports can be delayed before they change due to delay. They find research on this issue inconclusive and contend that the best timing for verbal reports has not been established and supported by research evidence. In studies using verbal reports, however, there have been two lines of argument concerning the timing of verbal reports. On the one hand, verbal reports should be as close as possible to the reading task to tap into the short-term memory, but on the other hand, the closer to the reading task they are (e.g. in prompted reports, or sentence-by-sentence reports), the more they may interrupt and potentially distort the natural flow and pace of the reading process. Delayed reports, while allowing for a less distorted reading process, may be marred by trace decay, intervening processes, and the reader's attempts to reconstruct the processes.

Pressley and Afflerbach (1995) disagree with Ericsson and Simon's (1984) contention that there is an important difference between participants' reporting and interpreting the content of their short-term memory. Ericsson and Simon argue against participants' explaining what they are doing because it induces them to make judgements about their processing. In contrast, Pressley and Afflerbach (1995) base their line of argument on metacognitive theory and claim that participants' explanations may reveal their greater cognitive awareness without impairing their processing. They speculate that if subjects explain their processing, this may provide insights into "sophisticated" processing. In their opinion, revealing certain processes outweighs the concerns that processing may be affected. However, they conclude that there is still not enough hard evidence illuminating differences between readers reporting the contents of their short-term memory or explaining their processes.

3.4.4 Instructions and think-aloud training of readers

Ericsson and Simon (1984) note two types of instructions for the readers. They can either encourage fairly general verbal reporting, or they can prompt the subjects to verbalize only specific processes such as predicting, asking questions, or inferencing. It is noteworthy that prompting specific types of verbalization cannot be representative of the natural processing and may bias the results. Ericsson and Simon also refer to various types of training or "warming-up" (1987, p. 37). They note that training should be standardized within a group of participants in order to make sure that the conditions for all participants are the same and the same verbal report procedure is used. They argue that subjects may not need special or elaborate training, but they may need reminding to resume thinking aloud during the task if they stop doing so.

With respect to the type of instructions — neutral vs. instructions used to elicit particular cognitive behaviour — Pressley and Afflerbach (1995) seem to share the view of Ericsson and Simon (1984). If a researcher prompts processing of a particular kind by asking subjects to attend to that process, the result may be affected. The instructions specified are bound to change the processing, as subjects are likely to attend to it more to please the researcher and this will inevitably bias the results. Consequently, Pressley and Afflerbach argue that the "researcher's silence about how the text might be processed is more defensible than directions that prompt particular processes especially when the goal is to learn about the processes people naturally use when they read" (1995, p. 133). It is worth noting, however, that a large number of empirical studies using think-alouds had particular research foci, such as inferencing or predicting, and specifically directed subjects to verbalize only along specific lines.

3.4.5 Completeness of method description

Pressley and Afflerbach (1995) argue that numerous studies do not give a detailed description of their methodology, making it difficult to evaluate and compare their conclusions. Their critical appraisal is that "some of the vagueness in reporting reflected the fuzziness in methods." (1995, p. 123). As a result, they suggest that studies should be as specific as possible in reporting the details about subjects, in particular their reading abilities and familiarity with the task, characteristics of the texts relative to the readers, directions given to the readers, type of practice before the task and, finally, the researcher's interventions during the task. The same level of specificity and completeness should be maintained in reporting methods of data analysis such as coding. Special attention should be paid to describing how the coding categories are developed, and how the protocols are coded, including an indication of inter-rater reliability.

3.4.6 Validity of think-aloud methods

Addressing the issue of the think-aloud method validity, Pressley and Afflerbach (1995) cite studies that found correlations between verbal reports and other measures (Olson et al., 1981, Meyers et al., 1990, Guthrie et al., 1991; Trabasso & Suh, 1993). Olson et al. (1981) found correlation between self-reported strategies and reading speed at particular points in the text. They measured reading times in experimental group thinking aloud and the control group reading silently. What they found was that both groups of subjects paused for thought or spent longer talking at the same places. The authors interpreted this as supporting evidence for the claim that there is a relationship between the processes going on during silent reading and reading while thinking aloud. Meyers et al. (1990) reported correlation between comprehension and self-reported strategies of college students. Guthrie et al. (1991) obtained a correlation between self-reported strategies and efficiency of search reading. Trabasso and Suh (1993) found correlations between self-reported inferencing, reading times, and recall. Even though they do not find much counter evidence (Wade et al., 1990), Pressley and Afflerbach (1993) argue there should be more validation of verbal protocols before any final firm conclusions are drawn about the validity of the method.

3.5 Strengths and limitations of think-aloud methodology

The think-aloud methodology has turned out to be an effective tool in a large body of empirical reading research in L1 and L2/FL. In a comparison of process- and product-oriented methodologies, Pritchard and O'Hara pointed to the key advantage of the think-aloud methodology: "From a research perspective, think-alouds provided insights into online processing behaviour that would have gone unrecognized in more typical, product-oriented data collection procedures" (2006, p. 157). Thus, in contrast to the quantitative data so often used in FL reading studies, the value of think-aloud data lies in the fact that they can provide an insight into hidden cognitive processes and tap into what would otherwise remain unnoticed.

Despite acknowledging the benefits of the think-aloud methodology, questions have been raised regarding its validity, especially with respect to veridicality and reactivity (Bowles, 2010). Veridicality refers to the degree to which think-aloud protocols truly and accurately reflect the subjects' processing due to the delay between reading and verbalizing. Reactivity refers to the effect that additional tasks (e.g. verbalizing) may have on readers' processing. In other words, it addresses the following questions. Does reading while verbalizing differ from silent reading? If yes, to what extent do readers' verbalizations differ from the actual process of silent reading? Lower-level processes, such as letter and word recognition, are fully automatized in skilled readers, and so they are inaccessible to subjects' conscious reporting. This may be different for higher-level processes if reading speed is slowed down by techniques that allow for concurrent verbalization, such as presenting texts sentence by sentence, or prompting the reader to stop and report at certain points.

In addition, think-aloud protocols may also depend on subjects' verbal skill. To minimize this effect, Ericsson and Simon (1993) suggested that subjects should be allowed to think aloud in their first or second language. Concerns about veridicality and reactivity of the think-aloud methodology as well as subjects' verbal skills have been scrutinized and tested in a number of empirical studies that have found support for the validity of the think-aloud method (Bowles, 2010; Leow & Morgan-Shrot, 2004).

3.6 Guidelines for think-aloud research

Theoretical and empirical investigations of think-alouds as a research tool have generated a series of practical guidelines for using think-alouds in order to maintain validity and reliability. These guidelines relate to the collection of data, especially with respect to timing and language of think-aloud protocols, the type of reported information, training of subjects and instructions, and the coding and analysis of the verbal protocols. Pressley and Afflerbach (1995) emphasised the key importance of three factors: maintaining the concurrency, encouraging reporting rather than interpreting or justifying, and neutral instructions without prompting particular cognitive behaviours. These guidelines have been taken on board in the research design of this study in order to maintain its validity and reliability.

3.7 Empirical studies using think-aloud methodology

To date, think-aloud empirical studies have generated ample data about the reading process. They have focused on cognitive processes and strategies used in reading, including readers' use of both linguistic and background knowledge. The studies have revealed the benefits and caveats of the think-aloud methodology. While the think-aloud methodology offers insights into elements of the reading process that would otherwise remain hidden and unobserved, certain criteria are necessary to assure the completeness, reliability, and validity of the think-aloud methodology. The concurrency of the think-aloud procedure is paramount. Readers should be given the choice to report in the language they find is easiest for them to verbalise their thoughts, either L1, L2 or a mix of both. The coding of verbal protocols needs to comply with data coding principles. The literature review suggests that the think-aloud methodology comprises both upsides and downsides. On the one hand it offers a view into cognitive processes and strategies used during reading, but on the other there may be a potential fear that the method could be affected by reactivity and veridicality. In other words, the literature suggests that the validity of think-aloud data could be marred. Thus, the study design needs to avoid these threats by assuring the concurrency of reporting, readers' choice of reporting language, and principled coding procedures. In addition, researchers ought to be aware of the fact that individuals differ in their ability to think aloud. Therefore, the research design of think-aloud studies should take account of these factors in order to reduce the think-aloud method effect.

This section provides a survey of empirical studies using think-alouds with respect to their research focus and validity. It gives a brief review of think-aloud studies that focused on:

1. Reading in L1,
2. Reading in L2,
3. The comparison of L1 and L2 reading.

3.7.1 Think-aloud studies of L1 reading

Think-aloud studies of L1 reading comprised think-aloud procedures that involved different types of tasks, texts types and their presentation. As regards the text presentation, they used both reading aloud and silent reading. Reading was either uninterrupted or interrupted. Pausing in interrupted reading was either done after every sentence, at certain points in the text marked by the researcher, or at the reader's choice. As shown in Table 3.1, the main research foci of think-aloud reading studies in L1 were:

a) The problem-solving process in L1 reading;
b) Strategies used in reading of different text types;
c) Strategies used by L1 readers and relating them to reading proficiency and readers' interest;
d) Differences between skilled and less-skilled L1 readers and training of strategies.

Table 3.1 Think-aloud studies of L1 reading

TA study	Text	Subjects	Research focus	Results
Olshavsky, 1977 Prompted TA method, retrospection	Short stories	24 US secondary school students	To identify reading strategies as regards readers' interest (high or low), reading proficiency (good or poor), and text type (abstract or concrete)	Greater interest, better reading proficiency, and abstract text are related to greater use of strategies. Readers' use of strategies depends on their interest, reading proficiency and text type.
Bråten & Strømsø, 2003 TA	Multiple expository texts	7 law students	To determine the strategies used and their change in a longitudinal study.	Readers' strategic processing changed over time. Different changes were observed for students at different levels of performance.
Kendeou & van den Broek, 2005 TA	3 expository texts	63 psychology students divided into 2 groups: readers with and without misconceptions	To identify the effect of readers' misconceptions on RC	Same type of processes used by readers with and without misconceptions. Readers with misconceptions had lower recall of information and fewer valid inferences.
Kaakinen & Hyönä, 2005 Combination of TA (sentence-by-sentence) and eye-tracking	1 expository text of unfamiliar content	36 L1 Finnish psychology students	To examine online RC of relevant and irrelevant text information	Relevant text information generated longer fixation time and better recall. Different RC processes were used for relevant text information: there was more paraphrasing of relevant information. Few explanatory inferences were generated, probably resulting from low BK. The validity of TA as measure of online RC was confirmed.
Cromley & Willis, 2016 TA	1 scientific text	24 undergraduate students of geology	To compare specific transition and flexibility with transitions between high- and low-gain readers as regards learning from text.	High-gain readers verbalized more vocabulary, BK, and strategies. They showed more flexible patterns before verbalizing strategies and vocabulary.
Jackson, 2016 TA pause protocols (at ends of paragraphs)	Science texts	48 students	To determine the effectiveness of TA to improve RC in scientific context.	TA increased readers' RC of a science text. It increased coherent representation of the text and improved self-monitoring.

Olshavsky (1977) conducted the first study by using think-aloud protocols to identify reading strategies. She wanted to find interactions between strategy use, reader's interest, reading proficiency level, and style of writing. She defined the variables operationally as: low and high interest, good and poor reading proficiency, and concrete and abstract writing style. Her first hypothesis was that readers with high interest would use strategies more frequently than those with low interest. She also assumed that good readers would not only use strategies more frequently than poor ones but would also use different types of strategies. Finally, she hypothesised that an abstract text would result in higher strategy use than a concrete text. Twenty-four students were asked to read a short story and verbalize their thoughts at the same time. Each subject read the story chosen according to their specific combination of interest, proficiency, and text style. The prompted think-aloud method was adapted from problem-solving research and piloting showed that it was more successful than letting subjects to verbalize whenever they felt like it. Each subject was given a training session so that students familiarized themselves with the task. The analysis of protocols revealed ten strategies, either word-related (use of context to define a word, synonym substitution, stated failure to understand a word), clause-related a (re-reading, inference, addition of information, personal identification, hypothesis, stated failure to understand a clause) and story-related (use of information about the story). All subjects were found to have used all strategies, yet some strategies were used more often by some readers. Three specific relationships emerged. First, greater interest was found related to greater use of strategies; second, better readers used strategies more; and finally, abstract text turned out to generate more strategy use. Olshavsky (1977) concluded that readers' use of strategies depends on their interest, reading proficiency, and writing style.

Bråten and Strǿmsǿ (2003) conducted a longitudinal study to observe how the use of strategies changes over time. They detected different changes at different levels of readers' performance. Strategy use was also observed as regards readers' reaction to different texts (Kaakinen & Hyönä, 2005) and readers' misconceptions (Kendeu & van den Broek, 2005). Specifically, readers' reactions to relevant and irrelevant information in texts was explored, and it was found that relevant text information generated longer fixation time and more paraphrasing. In addition to that, high-gain readers were found to show more flexibility when verbalizing strategies and vocabulary (Jackson, 2016). Think-aloud protocols were not only used as a research tool but also as an instruction tool to improve reading comprehension in scientific contexts (Jackson, 2016) and it was determined that using think-aloud protocols increased coherent text model monitoring. Finally, think-aloud studies of reading in L1 offered support for the validity of think-aloud methodology (Kaakinen & Hyönä, 2005).

Summary of think-aloud studies of L1 reading

Think-aloud studies conceptualized L1 reading as a problem-solving activity. When reading L1 texts (expository, descriptive, narrative), readers come across problems, both at a local and global level. To solve them, readers use a variety of strategies and knowledge sources, and link them to contextual information. Readers' text processing leads first to constructing a text model, followed by a situation model (Kintsch, 1998). Think-aloud studies of L1 reading provided data to identify strategy use of good and poor readers. Good readers were found to use more strategies and be more flexible in their application. They managed to adapt to the challenges posed by the text. In addition, they were found to monitor their reading, and link the text's content to their background knowledge. Strategy use can be taught, and so readers can improve their ability to identify comprehension problems and resolve them. Think-aloud studies of L1 reading also looked into the links between strategy use and readers' interest, reading skill, and the type of texts. If readers perceived texts as interesting, they tended to be reading more actively. Abstract texts were found more challenging for the readers' problem solving. In sum, think-aloud studies in L1 shed light on L1 reading. The think-aloud methodology has turned out to provide successful research tools. Its findings motivated parallel think-aloud research into L2 reading and a comparison of L1 and L2 reading.

3.7.2 Think-aloud studies of L2 reading

Think-aloud studies of L2 reading have predominantly interpreted L2 reading as a language issue, rather than a problem-solving activity as was the case in L1 reading. Table 3.2 presents studies that focused on the following factors L2 reading:

a) Successful and unsuccessful readers;
b) Readers with higher or lower L2 proficiency and higher or lower reading ability;
c) Specific strategies, especially lexical inferencing;
d) Validity of the think-aloud methodology.

Table 3.2 Think-aloud studies of L2 reading

Study **TA type**	**Text**	**Subjects**	**Focus**	**Results**
Hosenfeld, 1984 Concurrent TA, retrospection	Learning task	40 US students of L2	To compare skilled and unskilled L2 readers, to investigate the effect of reading instruction on readers' problem solving.	Differences observed in reading of skilled and unskilled readers. Instruction (awareness training) was effective.
Nassaji, 2003 TA, immediate retrospective report	1 passage	21 intermediate ESL students	Use of strategies and knowledge sources in L2 lexical inferencing.	Observed a relatively low rate of successful inferencing. The success of inferencing was related to the quality rather than quantity of strategies used.
Nassaji, 2004 TA, immediate retrospective report	1 passage	21 intermediate ESL students	To define the relationship between the depth of L2 vocabulary knowledge, use of lexical inferencing strategies, and success in determining the word meaning.	The depth of readers' vocabulary knowledge was found to be related to the use of certain strategies. It contributed to the success of the inferential process.
Leow & Morgan-Short, 2004 TA	1 text with 17 target forms	38 L2 learners of Spanish involved in the TA group and 39 in the non-TA group	To determine the effects of thinking aloud on readers' cognitive processing.	TAs did not have detrimental effect on RC. Reactivity was not detected.The translation strategy was used most frequently in both groups.
Bengeleil & Paribakht, 2004 TA	1 expository text	17 EFL medical students	To define the effect of L2 reading proficiency on L2 lexical inferencing.	To infer meanings of unfamiliar vocabulary, readers used a variety of knowledge sources, including L1 and BK. This applied to readers with high and lower L2 proficiency. More proficient readers were better at L2 inferencing.
Ghaith & Obeid, 2004 TA	1 passage	32 EFL students, divided into 2 groups: 1 receiving TA	To determine the effect of TA training on RC, both literal and higher-order RC.	Thinking aloud was positively correlated with RC, especially overall comprehension, critical comprehension, and interpretative comprehension.

Study TA type	Text	Subjects	Focus	Results
		instruction, and 1 not.		
Chang, 2006 TA concurrent, sentence-by-sentence	2 texts: on a familiar and an unfamiliar topic	55 EFL Chinese students	To determine the effects of topic familiarity and linguistic difficulty on the reading strategies and mental representations.	Readers mostly used local-level strategies and 2 global-level: monitoring and inferencing. They focused on linguistic problems, less on text structure. Unfamiliar text generated less inferencing, the difficult text more monitoring. Low L2 ability hindered the use of RC strategies.
Hu & Nassaji, 2012 TA	1 academic text from an economics textbook	11 ESL learners	To define the relation among the ease of L2 word inferencing, the type of inferential strategy and the retention of words.	A relationship was found between the strategy type and frequency of use and retention. The ease of inferencing was negatively related to retention. Retention depended on strategy type. Four categories of str. were used: form- and meaning-focused, evaluating, and monitoring.
Comer, 2012 TA	Informational texts	L2 students of Russian	To determine frequency and trends in lexical inferencing and other strategies.	A range of strategies was used. Rare use of explicit sentence- and paragraph-level context. BK used, but did not infer meaning of unknown words. Variation in strategy use inter- and intra-readers. Similar strategies as ESL readers but not as frequent and successful. The threshold hypothesis confirmed.
Kaivanpanah & Moghaddam, 2012 TA	3 expository texts	86 EFL learners	To define the role of reading proficiency and use of knowledge sources in L2 lexical inferencing.	Differences in L2 reading proficiency importantly affect inferencing success. Readers benefited from a variety of linguistic and non-linguistic knowledge sources and both local and global contextual cues. The success of lexical inferencing depended on how they combined them.

Study **TA type**	**Text**	**Subjects**	**Focus**	**Results**
Hu & Nassaji, 2014 TA	1 text from economics textbook	11 ESL students of economics and business	To define L2 inferencing of word meaning from context, and to identify the strategies readers use to do that.	Good inferencers showed the quality rather than quantity of strategy use. They used context, monitoring, sources of knowledge: linguistic and background. Poor inferencers used strategies in a random and unrelated way, relied on incongruent info, failed to resolve contradictions and integrate information.
Lee, 2014 TA, sentence by sentence technique	2 passages: 1 narrative, 1 informational	9 ESL primary school students	To identify patterns of inferencing when reading in English.	Readers showed limited use of bridging and global inferencing. They used more explanation in the narrative and less in the informational text. They used more paraphrasing, association and interpretation as compared to integration and text structure recognition.
Nalliveettil, 2014 TA immediate retrospection	1 semi-technical text	52 ESL engineering students	To assess reading strategies of successful and unsuccessful readers.	Successful readers used more strategies than unsuccessful readers. Successful readers also used prior knowledge and connected the text to their own experience.
Anvari & Farvardin, 2016 TA	3 passages	15 EFL students	To assess differences between successful and unsuccessful inferencers.	There were no differences found between more or less successful inferences as regards the number of strategies. They differed in the quality of use of lexical inferencing strategies.

Hosenfeld (1984) used verbal protocols in metacognitive awareness training of two L2 readers. The study was built on the assumptions that readers are able to monitor their reading performance and that good reading involves a discrete set of strategies to be used in awareness training of poor readers and thus serve as an instrument to improve their reading. Awareness training involved two steps: helping readers become aware of what strategies they used and comparing them to strategies commonly used by good readers. The researcher-instructor prompted the reader indirectly, only if necessary. The researcher then analysed the protocols and devised a profile of strategies to be used by the reader. Hosenfeld distinguished two groups of strategies: the first were strategies used in uninterrupted reading, and the second were used when the reader came across a problem, such as an unknown word. She divided the strategies into those used by good readers and poor readers. Good readers tended to keep the meaning of the passage in mind, read in broad phrases, skipped inessential words and guessed the meaning of unknown words from context. They had a good self-concept as readers. On the other hand, poor readers were found to lose the meaning of sentences immediately after decoding, and they read word-by-word, rarely skipping unknown words but turning to the glossary instead. Their self-concept as readers was not good. Hosenfeld then presented verbal protocols of remedial sessions with the researcher-instructor. She presented a list of tips used in remediation and showed the result of remedial sessions by profiling the readers' newly learned strategies.

Hosenfeld's study had pedagogical implications. It was argued that her training actually provided poor readers with strategies without teaching them language (Alderson & Urquhart, 1984), a claim that can hardly be substantiated after a close examination of the researcher's interventions in the remedial session with training primarily focused on problem solving strategies such as new vocabulary. In fact, she provided the learner with language in a less overt manner and only illustrated those interventions when the learner successfully picked up the trainer's prompts. In general, her work was concerned with mapping of reading strategies, especially in relation to word meanings and meaning retention while decoding.

Chang (2006) investigated how reading strategy use is influenced by readers' topic familiarity and linguistic difficulty. EFL Chinese students read a topic-familiar text and topic-unfamiliar text presented in the sentence by sentence think-aloud procedure.

Readers were found to use more local-level processing except for two types of global-level processing: monitoring and inferencing. They mostly focussed on linguistic problems, and this was at the expense of noting text structure and inter-sentential coherence. Two observations in relation to texts and strategy use were made. Topic-unfamiliar text generated less inferencing, mainly low level, whereas difficult text generated more monitoring strategies. The findings were interpreted to support the position that poor L2 ability hindered the use of reading strategies. Strategies were explored in a study by Nalliveettil (2014) that set out to assess reading strategies of engineering students, by using a semi-technical text. Successful and unsuccessful readers were compared with respect to their strategy use. Successful readers used more strategies than unsuccessful readers. They were observed to use prior knowledge and connected the text to their own experience.

A growing amount of attention in reading studies has been paid to lexical inferencing. Inferencing studies have explored the following issues:

a) Aspects and patterns of inferencing;
b) The relation of inferencing and knowledge sources, including L2 vocabulary;
c) The relation of inferencing and reading proficiency;
d) Good and poor inferencers.

In their study of various aspects of inferencing, Hu and Nassaji (2012) related the types of inference, the ease of inferencing and the retention of words. In their think-aloud study, they used one academic text from an economics textbook that was read by eleven ESL learners. They found a significant relationship between the strategy type, frequency of its use, and retention. The ease of inferencing was negatively related to retention. That is, if readers found it easy to infer a word's meaning, they were less likely to remember it. Retention depended on the type of strategy used. Strategies were grouped in four categories: form-focused, meaning-focused, evaluating, and monitoring. Patterns of inferencing were further investigated by Lee (2014). A narrative and an informational text were used in this think-aloud study, which aimed to identify patterns of inferencing when reading in English. Readers showed limited use of bridging and global inferencing. They used more explanation in the narrative text and less of it in the informational text. Overall, readers used more paraphrasing, association, and interpretation and less integration and text structure recognition.

A similar investigation into the frequency and patterns in strategy use was carried out by Comer (2012): English readers of Russian as L2 read informational texts. He found that readers used a broad range of strategies, but they rarely used explicit sentence- and paragraph-level context. They used their background knowledge, but rarely to infer meaning of unknown words. A great variation in strategy use was observed both between readers and also in think-alouds of individual readers. Although they used similar strategies in lexical inferencing as ESL readers, the frequency and success of strategy use was lower. This was interpreted as supporting the threshold hypothesis.

Another group of studies looked into the relation between inferencing and readers' knowledge sources, including L2 vocabulary. In his study, Nassaji (2003) investigated think-aloud and retrospective data of 21 intermediate ESL students. Overall, a relatively low rate of successful inferencing was observed. He argued that the success of inferencing was more related to the quality rather than quantity of strategies used. The prime importance of quality over quantity of strategy use was corroborated by Hu and Nassaji (2014) in their study of successful and unsuccessful inferencers. Nassaji (2004) added another variable and compared the relationship between the depth of L2 vocabulary knowledge, the use of lexical inferencing strategies, and the success in determining the word meaning. He found a relationship between the depth of readers' vocabulary knowledge and the use of certain strategies. Vocabulary knowledge contributed to the success of the inferential process. The finding that multiple sources of knowledge are used in L2 readers' inferencing process was corroborated by Bengeleil and Paribakht (2004) and Kaivanpanah and Moghaddam (2012). Readers benefited from a variety of linguistic and non-linguistic knowledge sources as well as both local and global contextual clues. The success of lexical inferencing depended on how readers combined knowledge sources and context clues. In contrast, Comer (2012) observed that although background knowledge was used in L2 reading, it was not used to infer meaning of unknown words.

Another key variable in think-aloud studies of inferencing was reading proficiency. Bengeleil and Paribakht (2004) aimed to explore the effect of L2 reading proficiency on L2 lexical inferencing. Seventeen EFL medical students read one expository text and did the think-aloud procedure. Readers with both higher and lower L2 proficiency used different knowledge, including L1 and background knowledge. Better L2 readers had more success in L2 inferencing. Similarly, Kaivanpanah and Moghaddam (2012) aimed to define the role of reading proficiency and use of knowledge sources in L2 lexical inferencing. They used three expository texts with 86 students and found that differences in L2 reading proficiency importantly affected inferencing success. Readers benefited from a variety of linguistic and non-linguistic knowledge sources and local and global contextual cues. The success of lexical inferencing depended on how they combined and coordinated them.

A group of think-aloud studies of L2 inferencing explored and profiled the differences and similarities between good and poor inferencers. Hu and Nassaji (2014) carried out a think-aloud study with 11 ESL students of economics and business who read one text from an economics textbook. Successful readers made use of context, monitoring, and various knowledge sources: linguistic, contextual, and background. They were distinguished by the quality rather than quantity of strategy use. Non-successful inferencers used strategies in a random and unrelated way. They relied on background knowledge incongruent with text information and could not resolve contradictions in the text. They lacked background knowledge and failed to fill in their gaps in text comprehension to integrate information. This finding was not supported by Anvari and Farvardin (2016), who aimed to assess differences between successful and unsuccessful inferencers in a think-aloud study involving 15 EFL students reading three passages. No differences were found as regards the number of strategies. They differed in the quality of use of lexical inferencing strategies.

Summary of think-aloud studies of L2 reading

Think-aloud reading studies have provided a conceptualization of L2 reading and generated extensive data about the differences between successful and unsuccessful readers in L2. The key factor in L2 reading is L2 proficiency and it was found to significantly affect the success of L2 reading. In exploring the role of L2 proficiency, think-aloud studies of L2 reading tested the existence of a threshold of L2 proficiency that allows the readers to fully use their reading skills and knowledge resources. The findings suggested a language threshold at which reading skills and strategies could be transferred to L2 (Comer, 2012).

In addition to L2 proficiency, the studies also explored the effect of reading skills and the use of strategies in L2 reading comprehension. The observation of strategy use in L2 reading indicated that L2 readers were more engaged in lower-level than higher-level processing. This underscores the claim that L2 reading can be primarily seen as a language issue, rather than a problem-solving issue and it distinguishes L2 reading from L1 reading. Still, both in L1 and L2 reading, metacognitive strategies and background knowledge were found to affect comprehension. For this reason, it has been suggested that L2 instruction should be paired with strategy training.

Think-aloud empirical studies of L2 reading provided ample evidence to identify the differences in strategy use between good and poor readers. Good readers were found to use a variety of strategies both at lower- and higher-level processing. They were L2 proficient, so they did not need to compensate for L2 proficiency. On the other hand, poor readers may have used their background knowledge, but failed to link it to the text, or they retrieved irrelevant background knowledge that was incongruent with textual information.

3.7.3 Think-aloud studies comparing L1 and L2 reading

Think-aloud studies comparing L1 and L2 reading focused on the following issues:

a) Comparison of strategy use in L1 and L2 reading;
b) Use of specific strategies (lexical inferencing and metacognitive strategies) in L1 and L2;
c) Problems encountered in L2 and L1 reading.

Table 3.3 Think-aloud studies comparing L2 and L1 reading

Study **TA type**	**Text**	**Subjects**	**Focus**	**Results**
Block, 1986 Concurrent TA, immediate retrospection	Texts from a psychology textbook	EFL college students	To compare strategy use of non-proficient readers in L1 and L2.	Readers' approaches to text were reflexive or extensive. Strategies used were general and local. Two strategy patterns identified were integrating and non-integrating. No differences were found in strategy use in L1 and L2.
Sarig, 1987 Concurrent TA	2 academic passages	10 EFL students, native Hebrew speakers	To identify strategy groups (termed "moves") and to compare them in L1 and L2.	4 types of strategy groups were identified. They were classified as promoting or deterring comprehension. Results were similar in L1 and L2 and prove transferability of strategies from L1 to L2.
Cavalcanti, 1987 TA pause protocols	2 introductions to academic articles in Portuguese and English	L1 Portuguese readers, L2/FL English readers	To identify problems encountered by L1 and L2/FL readers.	L2 and L1 readers used the same reading style. Faster reading and fewer pauses in L1 reading.
Davis & Bistodeau, 1993 TA, sentence-by-sentence pauses	2 newspaper articles in English and French	16 readers: L1 English and French	To identify and compare groups of reading strategies in L1 and L2.	In L1 English more top-down strategies were used. French readers used the same amount of strategies in L1 and L2. Strategies were seen as transferrable from L1 to L2 at a linguistic threshold.
Zwaan & Brown, 1996 TA	2 stories in L1 and L2	12 US students of French, skilled and less skilled L1 readers	To identify and compare groups of reading strategies in L1 and L2.	L2 reading depended more on L2 proficiency than L1 reading ability. L1 reading ability enhanced the use of context and dealing with unknown vocabulary in L2.
Jimenez et al., 1996 Prompted and unprompted TAs	Texts in English and Spanish	11 bilingual Latino and 3 English L1 students, divided into good and poor readers	To identify metacognitive strategies in L1 and L2.	Identified 3 groups of strategies: text-initiated, interactive and reader-initiated. Poor readers' main problems were unknown vocabulary and being unaware of the reading goal.

Study TA type	Text	Subjects	Focus	Results
Yamashita, 2002 TA, retrospective report	4 passages	17 L1 Japanese students of English, divided into four groups by reading ability in L1 and L2.	To identify and compare reading strategies in L1 and L2	35 strategies identified. Observed strategy transfer from L1 to L2 in all four groups, especially language independent strategies. L2 proficiency had a stronger effect on L2 reading than L1 reading ability.
Stevenson, Schoonen & Glopper, 2007 Concurrent TA	4 argumentative texts: 2 in Dutch and 2 in English	22 EFL Dutch students	To test inhibition and compensation; i.e. whether the lack of language affects processing, and whether it can be compensated by better reading skills.	Hypotheses refuted. Readers were not found to focus less on global text content in EFL. They did not compensate for language problems by focusing on global text content. 3 groups of strategies were identified: orientation (content, language), type (regulatory, cognitive), and linguistic processing domain (above, at, or below clause level).
Pritchard & O'Hara, 2006 Immediate retrospection	1 text from a science textbook in English and Spanish	20 Spanish bilingual students	To identify and compare strategies used by bilingual readers in L1 and L2.	In L1 readers used more inter-sentential strategies. In L2 they used more intra-sentential strategies. No strategy transfer between L1 and L2 was detected.

Block (1986) carried out a comparative study of strategy use in L1 and L2 by using a think-aloud method. The subjects were native and non-native college students who took remedial reading classes. Block chose poor readers with the assumption that they have not yet automatized their comprehension strategies and are therefore more aware of them, especially as regards solving problems encountered in reading. Block looked into the relationship between strategy use, memory, comprehension, and academic performance. The research instruments used apart from think-alouds were recall and a multiple-choice test. Verbal protocols were analysed and classified according to the "mode" of response (e.g. reflective and extensive) and strategy type (e.g. general and local). Ten general and four local strategies were identified, the distinction between them being that general comprehension strategies involve comprehending and the monitoring of comprehension, and local strategies involve understanding specific linguistic units. Further analysis of think-alouds showed two patterns of strategy use: integrating and non-integrating, based on the following criteria: the extent to which readers managed to recognize and integrate aspects of text structure, how much they used their personal experiences and associations, and if they responded in the extensive mode. Block did not find L1 to account for the differences in patterns of strategy use, as L2 readers did not use strategies or patterns of strategies different from those of L1 speakers of English. This led Block to conclude that strategy use is a stable phenomenon independent of language-specific features. In this respect Block's research was corroborated by further research evidence (Hudson, 1982), indicating that deployment of cognitive strategies does not depend on L2 proficiency but is readily transferable from L1 to L2. Finally, the analysis of recalls and comprehension test also showed that readers appeared to benefit from three sources: use of extensive mode, use of integration, and use of personal knowledge. Block pointed out that not only are there many commonalities in the strategy use of poor readers, but there are also many individual differences. For this reason, she avoided making any recommendations on how to deal with poor reading but argued for an individualized approach that draws on the think-aloud method as a consciousness raising and teaching tool.

Differences between strategy use in L1 and L2 were also explored by Sarig (1987). In a study involving ten Hebrew speaking students, she mapped four "move types", another term for a strategy group, and analysed the protocols as regards the effect of strategy use on comprehension. On that basis, strategies were classified as promoting or deterring comprehension. The results showed that comprehension promoting and deterring strategy use in the two languages was very similar. As to the level of comprehension, she further divided strategies as belonging to content proposition analysis, main idea selection, and overall message synthesis. Correlations between languages were high for three levels of comprehension promoting and deterring effects. Lower correlation was obtained only for content proposition as a comprehension promoting strategy. Sarig (1987) interpreted this result as showing that the lowest levels have the least ability for strategy transfer. The overall results of this study provided further evidence for the transferability of reading strategies from L1 to L2, with the differences being interpreted as arising from the differences in processing between individual readers rather than the differences between processing in L1 and L2.

Cavalcanti (1987) investigated the problems encountered by L1 and L2 readers when they needed to pragmatically interpret introductions to academic papers. She focused on the relationship of key words and potential and actual problems in pragmatic interpretation from the angle of the reader, researcher, and materials designer. Pause protocols were used, requiring the readers to read a text silently and think aloud at points when they naturally make a pause in reading. Cavalcanti assumed that when readers slow down their reading and go from quick automatic reading to a more controlled reading, this is likely to be caused by some problem they have encountered. Readers were found to have problems with thinking aloud and would read large parts of text without stopping. Their verbalizations were closer to retrospective accounts than concurrent think-alouds. The participants were specially trained by the researcher, but no specific information or detail is given about the nature of this training apart from remarking that it "led the subjects from retrospection proper to thinking aloud" (Cavalcanti, 1987, p. 239). Students read introductions both in their L1 (Portuguese) and L2 (English) and their pause protocols were triangulated with four measures: the title study task, interventionist procedure, oral summary, and selection of key vocabulary. Readers were found to use the same reading style in L1 and L2 but they read more quickly in L1 and made more pauses in L2. Cavalcanti found L1 protocols less informative than protocols in L2, so she speculated that the automaticity of L1 reading, which implies less controlled reading, may have contributed to that effect. She concluded that pause protocols are a good indirect route of investigating the reading processes, provided that they are triangulated with other measures and points of view.

Davis and Bistodeau (1993) carried a within-subject study of French and English L2 readers at undergraduate and postgraduate levels by using think-aloud methodology. After analysing the protocols, they identified three bottom-up strategies, eight top-down ones, and two metacognitive strategies. The comparison among readers showed that whereas English L1 readers used more top-down strategies in their L1 than in their L2, French L1 readers used a similar amount of different strategies in their L1 and L2. Because the L1 French readers were postgraduate students and English L1 readers were undergraduates, the results were interpreted not only as supporting the transferability of strategies from L1 to L2, but also as an indicator of a linguistic threshold which enables the transfer of strategies. Only beyond a certain level of proficiency were students able to use strategies in L2 like in L1.

Zwaan and Brown (1996) investigated the transfer of L1 reading ability to L2 reading by using think-aloud protocols in a group of twelve American university students of French who were in their second or third years of studying French and were classified as non-fluent. Half of the students were skilled in L1 reading and half were less skilled. They read two stories in L1 and L2 and their think-aloud protocols identified four groups of strategies: association, paraphrase, explanation, and prediction. They found a stronger effect of L1 ability on the use of strategies in L1 texts than L2 texts. Even though they were equally fluent in L2, skilled L1 readers were found to paraphrase and translate more accurately, and generally had less difficulty at the word level. The authors therefore argued that L1 reading ability enhances the use of context in deriving meaning of unknown vocabulary in L2. However, in comparing the effect of L2 proficiency and L1 reading ability on reading comprehension of the subjects in this study, the authors acknowledged that L2 reading depended more on L2 proficiency than L1 reading ability. The study can be seen as providing further evidence for the threshold hypothesis.

Jimenez et al. 1996 investigated metacognitive strategies used by fourteen bilingual Latina/o children of good and poor reading ability and three monolingual English children of good reading ability. Materials were piloted and used for prompted and unprompted think-aloud. The participants' prior knowledge was assessed using techniques developed by the Illinois State Board of Education (Pearson & Valencia, 1986 in Jimenez et al. 1996). This assessment involves extracting five key words from each text and asking students to write ten sentences of what they know about each. Background knowledge assessment was intended to help the authors anticipate possible difficulties for students in terms of vocabulary and inferencing. Unprompted think-aloud was used to obtain a "natural" account of student thinking, whereas the prompted one was accompanied by questions aimed at eliciting students' metacognitive strategies. Think-alouds were combined with free recall and interviews with each student. After transcribing and coding the protocols, they identified twenty-two different strategies, ten of which were high frequency ones. They classified them into three groups labelled as text-initiated, interactive, and reader-initiated. The main problem of both Latina/o groups was unknown vocabulary and successful readers generally resolved it by searching for cognates. The less successful Latina/o readers primarily appeared unaware of the goal of reading. This group of readers was also found to be less able to use their knowledge of Spanish to enhance their comprehension of English. The successful English readers, on the other hand, were not under the constraint of having to deal with unknown vocabulary and were found to use prior knowledge. The authors report that they appeared to be doing a different task than the bilingual readers, because they could focus more on comprehension. As to their prior knowledge, successful Latina/o readers used less strategy invoking prior knowledge while reading in Spanish than while reading in English. The implication drawn by the authors is that successful Latina/o readers should develop more awareness of the similarities between L1 and L2, as this would enable them to use the resources they have and develop an ability to assess the relative importance of individual unknown words. Moreover, they should be trained to deal with vocabulary essential for comprehension in three ways: by using context, invoking prior knowledge, and making inferences.

Yamashita (2002) carried out a study using think-aloud and retrospective reporting with 17 L1 Japanese students of English, divided into four groups by reading ability in L1 and L2. They read four passages and 35 strategies were identified. Strategy transfer from L1 to L2 was observed in all four groups, and this was especially strong for language independent strategies. L2 proficiency was found to have a stronger effect on L2 reading than L1 reading ability. Stevenson, Schoonen and Glopper (2007) set out to test the inhibition and the compensation hypotheses; that is, whether the lack of language affects processing, and it can be compensated by better reading skills. Twenty-two Dutch EFL students read four argumentative texts: two in Dutch and two in English. Both hypotheses were refuted. Readers were not found to focus less on global text content in EFL, nor did they compensate for language problems by focusing on global text content. Three groups of strategies were identified in relation to orientation (content vs. language), type (regulatory or cognitive), and linguistic domain of processing (above, at, or below clause level). Similarly, no strategy transfer between L1 and L2 was detected by Pritchard and O'Hara (2008). Twenty Spanish bilingual students read a text from a science textbook in English and Spanish and their strategy use in L1 and L2 was compared. Readers used more inter-sentential strategies in L1, whereas in their L2 they used more intra-sentential strategies.

Summary of think-aloud studies comparing L1 and L2 reading

This group of studies sought evidence to show similarities and differences between L1 and L2 reading, as well as the possibility of transfer of reading strategies from L1 to L2. Some think-aloud studies have detected evidence to support the view that there are similarities between L1 and L2 reading (Sarig, 1987; Cavalcanti, 1987). On the other hand, other studies have found specific differences between L1 and L2 reading: L1 reading generates more top-down strategies (Davis & Bistodeau, 1993), it is more automatized, faster and includes fewer pauses (Cavalcanti, 1987) and that L2 reading is more affected by readers' L2 proficiency than by L1 ability (Yamashita, 2002; Sarig, 1987; Pritchard & O'Hara, 2008). Metacognitive strategies drew special attention in comparative think-aloud studies of L1 and L2 reading. Some studies found that metacognition and monitoring comprehension characterized good readers in L1, but not in L2 (Pritchard & O'Hara, 2008). Conversely, other studies found a similar frequency of metacognitive strategy use in both L1 and L2 (Davis & Bistodeau, 1993).

The strategies surveyed in comparative studies of L1-L2 reading that used the think-aloud methodology were grouped with respect to the following criteria:

1. Level of processing: local or global strategies (Block, 1986); inter-sentential or intra-sentential strategies (Pritchard & O'Hara, 2008); top down or bottom up strategies (Davis & Bistodeau, 1993);
2. Context: integrating or non-integrating strategies (Block, 1986);
3. Effect on reading comprehension: strategies promoting or deterring RC (Sarig, 1987);
4. Use by good and poor readers in L1 and L2, individual/idiosyncratic (Hu & Nassaji, 2014; Anvari & Farvardin, 2016);
5. Factors influencing the deployment of strategies: L2 proficiency, L1 reading ability (Yamashita, 2002; Stevenson et al., 2007);
6. Orientation: content or language strategies (Stevenson et al., 2007);
7. Source: text-initiated, interactive, reader-initiated strategies (Jimenez, 1997).

The findings of comparative L1-L2 reading think-aloud studies led to two questions. The first refers to the question of transferability of L1 reading ability to L2 reading. The answers provided by these studies are not conclusive nor definitive. Some have found evidence in support of transferability (Block, 1987; Sarig, 1987; Yamashita, 2001; Davis & Bistodeau;1993), others have not (Pritchard & O'Hara, 2008). The second question addresses the possibility of compensation between the factors affecting reading comprehension. In other words, can better L1 strategies compensate for lack of L2 proficiency? No compensation effect was observed by Stevenson et al. (2007). Overall, think-aloud studies of reading in L1, L2, and comparative studies of L1-L2 reading have yielded a rich array of findings that affirmed the validity of think-aloud as a methodology, and provided better insights into reading process in L1 and L2.

3.8 Proposed methodological framework for present study and research question 7

The research design of this study is based on the guidelines derived from theoretical and empirical studies using the think-aloud methodology. This study uses the following theoretical principles for the think-aloud research: maintaining concurrency of think-aloud reporting, using prompted pause protocols, think-aloud reporting in readers' L1, and think-aloud training of subjects without specific instructions as to what they should report. Care was taken to follow the guidelines both in data collection and data analysis, so that method effect could be avoided and the validity of research would be assured.

The aims of this research are connected to the selected methodological framework and are formulated in the following question:

RQ 7: How do readers with high or low background knowledge read subject-specific texts?

Sub-question: What patterns/strategies do readers with high or low background knowledge use in the process of reading comprehension?

3.9 Chapter summary and conclusion

This chapter surveyed theoretical and empirical studies of reading using the think-aloud methodology. It profiled the issues that guided the research design of this study. The literature review explored three groups of think-aloud studies: studies in L1 reading, studies in L2 reading, and comparative studies of L1-L2 reading. Within each group it summarized the main trends and findings and presented the view of reading that they offered. This chapter also discussed the strengths and drawbacks of the think-aloud methodology with a particular focus on validity. Conclusions drawn from theoretical and empirical think-aloud studies of reading motivated the research design of this study. Research question 7 was formulated to address the qualitative aspects of the effects of background knowledge in L2 reading comprehension of subject-specific texts, and the answers are combined with the answers to the quantitative research questions presented in Chapter 2.

CHAPTER 4
THE STUDY

4.1 Chapter aim and overview

This chapter describes the study. It first outlines how the main variables were operationalized. It then presents research phases, including the development of the instruments (i.e. the grammar test, reading comprehension test, background knowledge test, post-reading questionnaire, and think-aloud procedure) as regards the participants and experts, the materials, the procedure, and the statistical analyses used. The chapter closes by considering the ethical aspects of the research.

4.2 Operationalization of the main variables and research phases

Chapters 2 and 3 discussed the theoretical grounds for the definition of the main constructs or variables in this study: the criterion variable Reading Comprehension (RC) and the three explanatory variables Background Knowledge (BK), Grammar Knowledge (GK), and Self-reported Knowledge (SRF). Reading Comprehension was operationalized as a score on a reading comprehension test of three subject-specific texts, explanatory variables Grammar Knowledge and Background Knowledge were operationalized as scores on a grammar test and background knowledge test, explanatory variable Self-reported Familiarity was operationalized as readers' assessment of their familiarity with the topic in a post-reading questionnaire.

As explained in Chapter 2, a number of studies operationalized background knowledge as either a) subjects' field of study (Alderson & Urquhart, 1983; Peretz & Shoham, 1990; Clapham, 1991; Ja'far 1992; Hill & Liu, 2012), or b) subjects' self-reported familiarity with a particular text topic (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Brantmeier, 2003; Lin, 2002; Nodoushan, 2003; Pulido, 2007; Leeser, 2007; Lee, 2007; McNeil, 2010; Eidswick, 2010; Rouhi & Ashgari, 2011; Lahuerta Martinez, 2013; Ashrafzadeh et al., 2015; Horiba & Fukaya, 2015). In line with questions raised about the consistency of operationalizations used in L2 reading studies (Cervetti & Wright, 2020; Jeon & Yamashita, 2020), this study was designed with two operationalizations of background knowledge: first, tested Background Knowledge (BK), and second, Self-reported Familiarity (SRF). It was expected this would allow a comparison their effect on the criterion variable and identification of any differences. A detailed operationalization of the four variables in the study (i.e. criterion variable: Reading Comprehension; predictor variables: Background Knowledge, Grammar Knowledge, and Self-reported Knowledge) is presented and discussed in Sections 4.1.1–4.1.3., followed by the development and validation of research instruments.

This study uses a mixed-methods design by combining quantitative and qualitative methods in testing and think-aloud studies. This approach was used to address the question of how much background knowledge influences reading comprehension and how the readers use their background knowledge in reading comprehension of subject-specific texts. Generally, mixed-methods are used to provide "a stronger understanding of the problem in question" (Creswell, 2014, p. 215) in contrast to using only one form of data. Loewen and Plonsky (2016, p. 118) argue that there are many benefits of collecting and integrating both types of data in a single study because quantification may involve a more systematic and objective approach, while qualitative methods may allow "studying phenomena in a way that preserves much of their natural state."

Table 4.1 presents research phases as well as the involvement of participants and experts. It lists a series of preliminary studies that included instrument development and trialling, divided into testing and think-aloud studies. The main study encompassed data collection and procedures in the testing and think-aloud studies. Finally, the last phase covers data analyses and results divided into the testing and think-aloud part.

Table 4.1 Research phases

RESEARCH PHASE			EXPERTS	STUDENTS
PRELIMINARY STUDIES	**Instrument development**	Grammar test: a priori validation	7 language experts	-
		Text selection	4 finance experts	-
		Reading comprehension test: item development	4 finance experts	-
		Background knowledge test: item development	4 finance experts	-
		Instructions for test administration	researcher	-
		Think-aloud prompts, training and instructions	researcher	-
	Testing study: Instrument trialling	Test administration: GK, BK and RC tests Post-reading questionnaire	researcher	20
	Think-aloud study: Instrument trialling	Test administration: GK, BK and RC tests Post-reading questionnaire Think-aloud procedure	researcher	2
MAIN STUDY: DATA COLLECTION & PROCEDURES	**Testing study**	Test administration: GK, BK and RC tests Post-reading questionnaire	7 language experts, researcher	358
	Think-aloud study	Test administration: GK, BK and RC tests Post-reading questionnaire Think-aloud procedure	researcher	24
DATA ANALYSES & RESULTS	**Testing study**	Test assessment: GK, BK and RC tests Validation Statistical analyses	researcher, 2 finance experts	-
	Think-aloud study	Coding of think-aloud protocols Validation Analyses	researcher, finance expert	-

4.2.1 Reading comprehension

Grabe (2009) defines reading as essentially a comprehension process which entails cognitive processing with the ultimate goal of comprehension (2009, p. 14). Comprehension can be considered either from the process angle or the product angle; the former attempts to explain the nature, processes, and strategies of reading, whereas the latter focuses on the outcomes or results of reading. This study combines the product and process approaches to reading comprehension by using testing methodology and think-aloud methodology.

The testing methodology is focused on the outcomes of comprehension. Several frameworks for constructing reading comprehension tests have been proposed (Weir, 1993; Alderson, Clapham & Wall, 1995; Bachman & Palmer, 1996; Alderson, 2000). They specify the relationships among the reader, the text, and the task in sets of specifications for reading comprehension tests. They comprise test conditions and test operations. Test conditions refer to the circumstances under which operations are performed: the purpose of reading, the nature of the text, the presentation channel, the text length, the processing speed required, and the response mode. As for the reading comprehension test conditions in the present study, Section 4.5 outlines the following aspects related to texts and tests:

a) Text-related conditions: purpose, text type, text length, text organization, text topic, vocabulary, channel of presentation;
b) Test-related conditions: test structure, tasks, order of questions, processing speed, test format, item types.

Test operations are the skills readers are using in the performance of an EAP reading task, such as reading carefully for main ideas and important detail, or reading expeditiously for specific information (i.e. scanning) or gist (i.e. skimming). The choice of reading comprehension test operations in this study is reading carefully for the main ideas and important detail that covers understanding explicitly stated ideas, as well as those that are expressed implicitly. To understand implicit ideas, readers need to make inferences and use their background knowledge (Grabe, 2009). The operations chosen in this study involve reading both at a global and local level, and readers' ability to infer the meaning of macro propositions, lexical items, and pronominal references. Reading carefully for main ideas and supporting detail is not only the most common type of operation in reading comprehension test specifications, but it is also a central operation in theoretical models of reading (Kintsch & van Dijk, 1978; Rayner et al., 2011).

The think-aloud (TA) methodology is designed to tap into the reading process and strategies readers use. In this study it was used to shed light on how readers use their background knowledge when reading subject-specific texts. The variation used was prompted (pause) protocols (Cavalcanti, 1987) requiring subjects to verbalize at points marked in the text. This type of thinking aloud ensures that subjects report their thoughts concurrently while reading. It allows for the concurrency of reporting, but it does not remove the limitation of thinking aloud because it inevitably interrupts the reading process. The theory underlying the think-aloud methodology and the results of empirical studies using it were discussed in Chapter 3. Two operationalizations of reading comprehension outcomes and process were used in this study:

1) Score on a reading comprehension test based on four subject-specific texts,
2) Think-aloud protocols taken during reading comprehension.

4.2.2 Background knowledge

Koda (2005) links reading comprehension and background knowledge by arguing that: "Comprehension occurs when the reader extracts and integrates information from the text and combines it with what is already known" (2005, p. 4). What is already known, commonly referred to as background knowledge, has been traditionally operationalized as the subject's field of study or assumed experience, the subject's self-assessed familiarity with the topic, their free introspection, or an interview. Sometimes a combination of different background knowledge measures was used, including familiarity assessment paired up with reading habits, main study discipline, interests, or self-reported knowledge. As discussed in detail in Chapter 2, the measures of background knowledge used in most empirical studies were indirect, subjective, and not matched to specific reading texts. Direct measurement of background knowledge was generally avoided, sometimes due to concerns over confounding variables (Spyridakis & Wenger, 1991).

Overall, operationalization of background knowledge in reading studies may be problematic for several reasons. If background knowledge is equated with a study discipline or belonging to a culture, this overlooks the fact that people may have certain knowledge even if they do not have an official background in a discipline or do not belong to a specific culture. Thus, the assumption that one's study discipline or belonging to specific culture means that the person has certain background knowledge may be incorrect, or at least a very crude and subjective measure of background knowledge. People outside certain disciplines (or cultures) may possess particular discipline- or culture-related background knowledge because of their interests, experience, or other reasons. In addition, it is misleading to assume that all people within a certain discipline have the same amount of background knowledge, as there is most likely variation in the level of knowledge. In a similar vein, studies that used free introspection to assess background knowledge may have failed to elicit specific facets of knowledge not because subjects lacked it, but because they may have simply not reported it. Free introspection may offer only a very rough indication of one's background knowledge. Thus, it cannot be taken as a precise and specific measure of subjects' background knowledge.

In this study, background knowledge was operationalized in two ways:

1) A score on a background knowledge test (BK);
2) Self-assessment of readers' familiarity with the topic in a post-reading questionnaire referred to as: Self-reported Familiarity (SRF).

Stemming from the critical assessment of background knowledge studies, the background knowledge test in this study was developed by subject experts and was designed to measure those aspects of discipline-related background knowledge that the finance experts considered essential for the comprehension of the subject-specific texts used in this study (See Section 4.6). The task format was short answer questions. Students first completed the background knowledge test, then they proceeded with reading the first text and doing the corresponding reading comprehension test, followed by the assessments in the post-reading questionnaire. The same procedure was repeated for each text.

4.2.3 Grammar knowledge

Grammar knowledge was operationalized through a score on a grammar test based on a modified and validated version of the TEEP grammar test (Appendix A-1). Two main issues were addressed: how to minimize the effect of reading in the grammar test, and how to reduce the effect of testing lexis in the grammar test. Neither problem can be completely avoided; however, both were minimized by the selection of task type and validation of the items by a panel of language experts. The construction and validation of the grammar test are described in detail in Section 4.5.

4.3 Students and experts

The present study was designed with three groups: students, subject experts, and language experts.

4.3.1 The sample: students

In the preliminary and main studies a total of 404 students were involved, all from the School of Economics and Business, University of Ljubljana (Slovenia), from 21 to 22 years of age. The preliminary study involved 20 students (six males and 14 females) that trialled out all the instruments and two students that trialled the think-aloud procedure. In the main study 382 students (142 males and 242 females) were involved. Of these, 358 students took part in the quantitative (testing) study, and 24 students participated in the qualitative (think-aloud) study. The sample was homogeneous in the sense that it consisted of non-native speakers of English, all second- and third-year economics students, all of whom had the same L1 background: Slovene. The results of 14 international students who also took part in the study were not used in the analyses in order to keep the sample homogeneous regarding students' L1. The sample's homogeneity with regard to participants' L1 meant that it could be claimed that the potential effect arising from differences in L1 was eliminated. All students in the sample had learned English as an L2 for six years in primary and secondary school. At the School of Economics and Business they were taking two courses of English for Economics and Business and using the textbook *Market Leader Upper Intermediate, New Edition* (Cotton, Falvey & Kent, 2001). The students' level of proficiency in English could be assessed approximately as B2 of the Common European Framework of Reference for Languages (2001).

4.3.2 Finance experts

The focus on highly specialized background knowledge and the aim of maintaining ecological validity necessitated the involvement of subject experts at all stages of the preliminary and main studies, more precisely in: 1) the text selection, 2) the reading comprehension test design and item writing, 3) the coding, assessment, and validation of answers as part of the analysis of the results, and 4) the interpretation of findings. The choice of financial experts as subject specialist informants (Douglas, 2000) and raters in the study is in line with Pill and Smart's (2020, p. 137) assessment that expertise in the domain is a key fixed rater characteristic. Seven finance experts were involved in the text selection, three in the test design and item writing, and two in the coding, assessment, and validation of answers. All subject experts participating in the study held PhDs in finance, of whom two obtained their degrees in the UK and Austria. One expert was a full professor, three were associate professors, and three were assistant professors of finance. Three experts had worked as visiting lecturers at British, Belgian, Dutch, and Austrian universities. The number of subject experts involved in the study varied at different stages because of their busy teaching schedules and research commitments.

4.3.3 Language experts

Seven language experts were involved in the validation procedure of the grammar test. They focused on the construct validity; that is, whether the test measures what it purports to measure. Apart from taking part in the preliminary study, language specialists were also instrumental in the administration of the test battery in the main study. All language experts had taught at the School of Economics and Business for more than five years and were familiar with the students and the teaching situation. Two of them held a PhD, one in lexicography and the other in English literature. One held an MA in linguistics; four were studying for their MA degrees—two in TEFL, one in linguistics, and one in business administration.

4.4 Preliminary studies 1 and 2

This section describes the development, trialling, statistical analyses, and subsequent modifications of the instruments that were used in the main study. The sequence is in accord with Green's (2020, p. 116) claim that trialling is an essential element of quality control that precedes test analysis and modifications. The following set of instruments was developed:

1. The grammar test;
2. The background knowledge test based on four texts;
3. The reading comprehension test based on four texts;
4. The post-reading questionnaire;
5. The think-aloud procedure.

The specific objectives of the trialling, statistical analyses, and subsequent modifications of the instruments were:

a) Construct validation: making sure that the test measures what it purports to measure.
b) Increasing internal reliability of the test by eliminating the items with poor discrimination.
c) Improving the practicality of the test battery by appropriate timing of the test administration and scaling down the battery to a manageable length. The reason for this was that a sample of over 300 students was planned for in the main study. It was clear that it would be difficult to obtain such a large sample of students outside their class time, so it was decided that the study would be integrated into their course time. As a consequence, this meant that the administration of the battery of instruments could not exceed 90 minutes, a common lecture length at the university. Furthermore, a limit of 90 minutes also seemed reasonable because of the fatigue factor. It would be very unrealistic to expect students to participate longer than that without any effect on their performance. Finally, due to its length, the think-aloud trialling was scheduled during out-of-class time.

The trialling of the five instruments is described in detail in Sections 4.5, 4.6, and 4.7 with regard to participants, materials, procedure, analyses, and modifications of the instruments that were used in the main study.

4.5 The grammar test

The grammar test used in the preliminary studies was part four of the TEEP test battery (Test of English for Educational Purposes), also referred to as the Test of English for Academic Purposes (TEAP) (Appendix A-1). The TEEP was developed by Weir (1983) at the Centre for Applied Language Studies of the University of Reading in order to provide universities with a comprehensive picture of English language proficiency of students for whom English was their L2. The TEEP was considered as a "test which provides a measure of prospective students' ability to cope with their intended course of study" (Green, 2000). In its history, the TEEP has been revised several times in order to reflect the current models of language competence, respond to the needs of overseas students that intend to study at British universities, and provide a fair picture of candidates' English language proficiency in an academic setting. After major revisions by O'Sullivan and Slaght from 2000-2001, the new TEEP offered a more reliable and fairer assessment of test-takers' proficiency in reading, listening, and writing.

The TEEP grammar test used in Preliminary Study 1 contained 60 multiple-choice items with a reliability figure of 0.92 (Weir, 1983). The test was validated and calibrated to the target population of this study. The rationale behind the validation procedure was threefold:

1. Screening test items with regard to their construct validity;
2. Screening test items for cultural and historical bias;
3. Screening test items based on the results of statistical analysis of a pre-trial sample.

Triangulation of the information above was expected to give enough evidence to filter out those items that do not perform well on the target population, reduce the test to 30 validated items, and modify any items if necessary. The grammar test validation procedure was carried out in two preliminary studies described in the following sections.

4.5.1 Construct validation of the grammar test: Preliminary study 1

1. Aim

The main aim of Preliminary Study 1 was construct validation of the instrument for measuring grammar knowledge. My exploratory contact with the TEEP test developers at the Centre for Applied Language Studies at the University of Reading clearly indicated that the original TEEP test was being remade with observations and aims similar to my own (Slaght J., personal communication). They found some items in the grammar test to be more lexis-oriented than grammar-oriented, and their remaking of the test was aimed at improving the construct validity of the test. TEEP was also used by Shiotsu (2010) and underwent construct validation, the procedure necessary to validate instruments for their intended purposes, populations and contexts (Révész & Brunfaut, 2020, p. 30).

2. Method

To examine the construct validity and potential historical and cultural bias of test items in the grammar test, a rating instrument was designed. For each of 60 grammar test items it contained three questions: how the test item measures the knowledge of grammar, whether the item is culturally biased, and whether the item is historically biased. To answer the first question, the experts were offered a five-point Likert scale, whereas the questions on cultural or historical bias were dichotomous (see Appendices A-2 and A-3). In addition, experts were invited to comment on their choices where they felt it necessary. The language experts' assessments, as the first step in the validation procedure, focused on the construct validity; that is, whether the test measures what it purports to measure. Raters were given a booklet that contained all 60 test items printed on the left side, and the questions for their assessment of each item on the right side. This layout was chosen to prevent any possible mixups that could have occurred if raters had been given separate assessment sheets.

3. Language experts

As noted in Section 4.3.3, seven language experts were involved in the validation procedure of the grammar test. The expert who holds a PhD in lexicography was consulted again in the final selection of those items that measure grammar best and do not confound the measurement of the variable with a lexical element.

4. Procedure

Language experts were contacted directly and invited to participate in the study. They were given the booklets with the grammar test and questions. All seven language experts completed their ratings and returned the booklets to me. As it was clear that the main issue in construct validation was the distinction of purely grammatical items from the lexical ones, the expert on lexicography was consulted to give her view on the choice of items. Finally, the raters' and lexicographer's opinions were combined and the final choice of test items for the preliminary trial was made.

5. Changes to the grammar test in Preliminary study 1

Appendices A-2 and A-3 present language experts' ratings for each of the 60 items in the original grammar test. Appendix A-2 shows the results of the assessment of items for their appropriateness for measurement of grammar, whereas Appendix A-3 lists the results of filtering out items with cultural and historical bias. The expert assessment clearly indicated that there were grounds for screening the test for items that did not appear to measure grammar. Altogether, 18 items were rated as measuring grammar "very badly" by at least one judge (items 5, 8, 12, 17, 18, 19, 20, 23, 24, 29, 41, 42, 44, 47, 48, 54, 58, 59), and 29 items were found to measure it "badly" by at least one judge (items 1, 6, 7, 8, 9, 10, 11, 12, 17, 18, 23, 27, 28, 29, 30, 31, 32, 33, 35, 38, 42, 44, 47, 48, 49, 54, 57, 59, 60) (Appendix A-2). The ratings that pointed to the problematic items were also supported by the experts' comments (Appendix A-3). For instance, item 8 obtained the lowest rating of all items, with five raters judging it to measure grammar "very badly" and one "badly".

8. The ..... industry is very important at the present time to our economy.
A) oil B) foreign C) modern D) light

Raters' comments on this item expressed very clearly their opinion that the item tests lexical rather than grammatical knowledge:

- Rater A: This is a lexical item.
- Rater B: This item doesn't measure grammar.

- Rater C: Checking lexical knowledge.
- Rater D: Not grammar but vocabulary.
- Rater E: What are we checking here: lexical phrase, use of the article?

Items 18 and 54 were found to be testing grammar “very badly” and “badly” by four experts:

18. This exercise ..... us with a number of interesting possibilities.
A) sets B) shows C) gives D) presents

54. The university was ..... to those who had difficulty paying their fees.
A) pitiful B) hopeful C) sympathetic D) punishable

Items 29, 42 and 48 were rated as “very bad” and “bad” measures of grammar by three raters.

29. The library ..... nearly two million volumes.
A) consists B) compares C) composes D) contains

42. The department was ..... to his application.
A) unaware B) opposed C) contrary D) uncaring

48. "I am taller than you ..... three inches."
A) with B) by C) of D) in

Then there were nine items that were assessed as “bad” and “very bad” measures of grammar knowledge by two judges, and 21 items that were rated as “bad” or “very bad” measures of grammar by one rater. The raters not only pointed to the items that in their opinion clearly belong to lexis, but also detected items that could be categorized as “grammatical or syntactic collocations” (Benson, 1985; Carter, 1998; Schmitt, 2000), typically a combination of verb, noun, or adjective plus a preposition. These collocations are generally treated as borderline items between lexis and grammar. Item 24 illustrates this point and experts’ comments are given below:

24. P.T.O. stands ..... "Please Turn Over".
A) as B) like C) for D) by

- Raters A & D: Grammatical collocation.
- Rater B: Grammatical or lexical knowledge?

Apart from judging how items measure grammar, the raters also provided their assessment of items for cultural and historical bias. This generated an overall majority of ratings finding the items culturally and historically unbiased and therefore acceptable (Appendix A-3). Nevertheless, some raters observed that test items mostly had male subjects and were relevant to the British context. The following items were found to be culturally biased: 8, 20, 25, 42, 47, and 48; whereas items 8, 20, and 42 were considered historically biased. What is noteworthy here is that almost all of these ratings were given by one rater, whereas items 8 and 20 were flagged by two raters. Both items 8 and 20 were flagged as inappropriate on all three counts: not measuring grammar and historical and cultural bias:

8. The ..... industry is very important at the present time to our economy.
A) oil B) foreign C) modern D) light

20. "John's very friendly. He's from ..... England."
A) the north B) the north of C) north the D) north of the

The rating of these items was also reflected by raters' comments to item 20:

- Rater E: Wouldn't John from the South of England be kind? Geographic and historical bias.
- Rater D: Why not Hans from Germany?

When all experts' ratings were tabulated, an interview was held with the expert who holds a PhD in lexicography, and the expert who did his MA on collocations. All the items marked as a "bad" or "very bad" measure of grammar were compared and a final list of test items was agreed on. It was obvious that an item could be rated as a bad or very bad measure for various reasons, but the main problem with most of the items deemed bad or very bad was actually the overlap with vocabulary. The lexicographer stressed that grammar and lexis are inextricably linked and sometimes it is difficult to say that an item only measures grammar. She emphasized that how students actually learn grammatical collocations should also be considered. In her view, they are learned the same way as lexis. A similar observation was made by Shiotsu (2010, p. 46), who also used TEEP to test L2 syntactic knowledge and stated:

> "it is in fact simply impossible to remove the effects of lexical knowledge if one is to measure someone's syntactic knowledge, since it presupposes at least some awareness of word categories which depends on knowledge of word forms a part of the lexical knowledge."

The language experts' judgement, including the lexicographer's assessment of all items, was used as a basis for a principled selection of items. The following items were filtered out to reduce the original TEEP grammar test:

1. Items rated by experts as measuring grammar "very badly": 5, 8, 12, 17, 18, 19, 20, 23, 24, 29, 41, 42, 44, 47, 48, 54, 58, and 59.
2. Items deemed as purely lexical items by the lexicographer: items 5, 8, 19, 20, 29, 30, 32, 37, 41, 47, 53, 54, and 58.
3. Borderline grammar-lexical items, including grammatical collocations, as deemed by the lexicographer: 11, 12, 17, 18, 23, 24, 35, 42, 44, 46, 48, 59, and 60.

In sum, an a priori validation of the test items was carried out to improve the construct validity of the grammar test. The aim was to triangulate the results with those of a pre-trial of the test on a sample population of 20 students and in this way yield the final set of items to be kept in the revised grammar test for the main study.

4.5.2 Grammar test pre-trial: Preliminary study 2

1. Aim

Preliminary Study 2 set out to test how the grammar test performed on a sample of the target population. It was designed to improve the test's practicality by reducing the number of items in the test, while maintaining acceptable reliability figures for the instrument. It was felt that if the number of test items could be reduced from 60 to about 30, this would shorten the test administration, and make it easier to administer the whole battery of five instruments to be used in the main study in under 90 minutes.

2. Method

The final version of the grammar test to be used in the main study was constructed on the basis of two filters: the construct validation conducted in Preliminary Study 1, and acceptable reliability figures of the instrument trialled in Preliminary Study 2. This was followed by three steps:

1. Pre-trial administration of the grammar test to a population of 20 students.
2. Statistical analysis of the pre-trial on a sample of 20 students comprising the following data:
    a) Facility values of test items or the percentage of test-takers answering the item correctly.
    b) Contribution of a particular item to the reliability and internal consistency of the instrument as expressed by Cronbach's alpha when a particular item is removed from the test.
3. Modification of the test by eliminating the non-performing items and modification of items based on teachers' assessments and pre-trial results.

The three-step procedure made it possible to reduce the TEEP grammar test from 60 items to 30 items. It was expected that the selected items would perform well in the main trial and also allow a reduction in the test completion time from 30 minutes to 20 minutes. The aim was to limit the fatigue factor while still maintaining the validity and reliability of the instrument. This seemed particularly important, because the students were expected to continue on to the reading comprehension tests after completing this test.

3. Participants

Twenty students (six males and 14 females) took part in Preliminary Study 2. They were administered the whole battery of instruments, including the grammar test with 60 items, the background knowledge test, a reading comprehension test based on four texts, and a post-reading questionnaire. Two students also trialled the TA procedure.

4. Procedure

The test battery was administered to students outside of their course time. The grammar test was the first instrument of the battery, followed by the background knowledge test and the reading comprehension test. The administration of the test battery took a total of 110 minutes: 20 for the grammar test, 20 for the background knowledge test, and 70 for the reading comprehension test. (Table 4.8). The time was not fixed in advance and the students were allowed to write as long as they needed. They were also told that one of the aims of this trial administration was also to set the appropriate time limits for each section of the test battery. Having administered the test battery myself, I observed that the pacing of individual test-takers was similar.

5. Changes to the grammar test in Preliminary study 2

Appendix A-6 summarizes the statistical analysis of the grammar test. It gives the following parameters: facility values or share of correct answers for each question, and reliability of items measured by Cronbach’s alpha for all 60 items and Cronbach’s alpha if an item is deleted. Comparing these values offers a way to assess how difficult an item is and whether it contributes to the internal consistency of the test and reliability.

Facility values for the 60 items ranged from 5% for one item (item 3) to 100% for five items (1, 2, 10, 13, and 17). There were ten items with 95% correct answers, and ten items with 90%. At the lower end, apart from one item with 5%, there was one item with 20% correct answers. The overall distribution of results indicated that the students found 25 items to be relatively easy, as the number of correct answers was between 90 and 100 percent. Items that were too easy for most test-takers cannot discriminate among them and fail to offer information about them. This suggested that these items should be eliminated, and those that discriminate better should be kept as shown in Appendix A-6. This is consistent with Davies et al. (1999), stating that the majority of items should not be too difficult or too easy for the test-takers.

Table 4.2 Grammar test: facility values

FV (%)	100	95	90	85	80	75	70	65	60	55	50	45	40	30	20	5
N of items	5	10	10	5	4	4	3	2	4	4	2	2	2	1	1	1

The reliability of the grammar test indicated by the Cronbach's alpha result was 0.894 for all 60 items. Several tests were run to obtain Cronbach's alphas if the following eliminations were made:

1. All lexical items (as listed in Preliminary Study 1);
2. All borderline items (as listed in Preliminary Study 1);
3. All items with facility values above 90%.

The target was to reduce the test items from 60 to about 30; however, because of the relatively high number of items with facility values over 90%, it was clear that all three filters could not be applied. The number of items that would remain would be too low, which would in itself reduce the reliability indicated by Cronbach's alpha. A decision had to be made regarding which filter was more important: either the screening of items for their construct validity that would be basis for elimination of lexical and borderline items, or the screening of items for difficulty as indicated by facility values of over 90%. Following the premise that construct validity is "the basis on which the other approaches rest" (Cronbach, 1988, p. 151), the decision was made in favour of construct validity, so the first two rounds of omissions were 12 lexical items and 14 borderline lexical-grammatical items, followed by four items with facility values of 100%. This solution was not ideal, but it was the best solution at hand.

The outcome of both preliminary studies was the development of a modified grammar test to be used in the main study that had the following characteristics:

1. The instrument was validated in terms of its construct (Preliminary Study 1) by eliminating lexical items.
2. The instrument was tested for its reliability as shown by the test total Cronbach's alpha of 0.801, which is considered a relatively high degree of reliability in L2 research (Plonsky & Oswald, 2014).
3. The practicality of test administration was improved by reducing the number of items to 30.

4. The administration of the instrument was timed on the target population, so that the time length of administration was adjusted for the main trial.

The central issue that arose in the development and validation of the grammar test was how to balance validity, reliability, and practicality. Traditionally, validity referred to the fact whether the test tests what it is supposed to test, reliability related to the consistency of measurement, and practicality to the ease of test administration. In their review of validity conceptions, Chapelle and Lee (2022) list multiple concepts that have evolved in language testing since 1960's. Most of the early approaches were primarily focused on the quality of the test, such as whether the test samples the content adequately (i.e. content validity), how the test users perceive the test (i.e. face validity), and how well the test reflects the theoretical assumptions about the construct being measured (i.e. construct validity), whether the scores can be related to other measures (i.e. criterion-related validity), whether the test is equivalent to other tests (i.e. concurrent validity) and whether the test score can predict future achievement (i.e. predictive validity). Chapelle and Lee (2022) contend that newer and current views of validity that developed following Messick's (1989) work moved on from focussing on the test quality to "a more complex investigation of multiple dimensions of score meaning and use" (Chapelle and Lee, 2022, p.18). According to Brunfaut (2022), these trends in validity research resulted in the development of more practical validation frameworks, including two most influential ones: the argument-based approach (Kane, 1992) and the socio-cognitive approach (Weir, 2005).

The argument-based approach was proposed and refined by Kane (1992, 2006, 2013, 2022). This validation framework conceptualizes validity as building coherent arguments that back up interpretations and uses of scores. The validity arguments are structured through claims one would like to make about test scores. The validity arguments specify various elements of score meaning and lead to formulating claims about the utility of the test scores for particular purposes.

The socio-cognitive approach to validity and validation (Weir, 2005), which was adopted in this study, offers "the first systematic attempt to incorporate the social, cognitive, and evaluative (scoring) dimensions of language use into test development and validation" (O'Sullivan and Weir, 2011, p. 20). This framework links test takers, their responses, and test scores on the one hand with cognitive validity, context validity, scoring validity, consequential validity, and criterion-related validity on the other. Each of these validity aspects is associated with specific questions and their answers may serve as validity arguments to stakeholders. For example, in relation to cognitive validity the suggested questions are:

> "Are the cognitive processes required to complete the task appropriate? Are candidates likely to use the same cognitive processes as they would if performing the task in a 'real world' context?" (O'Sullivan and Weir, 2011, p. 21)

Overall, the socio-cognitive approach to test development and validation has the test-taker at the heart of the process, requiring the test-developer in particular to recognize the mental processes of the test-taker while doing a test task. The original socio-cognitive framework model (Weir, 2005) was refined in several steps (O'Sullivan and Weir, 2002, 2011; O'Sullivan, 2011, 2014, 2016) so that the latest version comprises three main elements: the test taker, the test system, and the scoring system. Chalhoub-Deville and O'Sullivan (2020) draw parallels and establish a clear relationship between the socio-cognitive model of validity and validation and Messick's (1989) six validation criteria.

Validation was a key concern in this study both as regards individual test items and the test as a whole. Construct validity was considered paramount and a rigorous a priori validation of the grammar test was undertaken by engaging language experts, one of them a specialist in lexicography, so that all items deemed not to be a measure of grammar were omitted as well as most borderline items that were categorized as grammatical collocations (Benson, 1985; Carter, 1998; Schmitt, 2000). Regarding the reliability of an instrument, Oppenheim states that it is "a precondition to validity" and "means consistency", but "is never perfect; it is always a matter of degree" (1992, p. 159). With this view in mind, the reliability and internal consistency of the test were measured and the results pointed to a relatively high level of reliability of the instrument (Cronbach's alpha .801). The final consideration taken into account was the test's practicality, which commonly refers to the time taken to construct a test, administer it, score it, and interpret its results. In this study, practicality primarily referred to the time necessary to administer the test. The outcomes of the construct validation as well as the reliability and practicality check were used as a filter to reduce the number of test items and develop a modified version of the grammar test that was used in the main study.

4.5.3 Reliability of the final measurement instrument (main study)

In order to establish the reliability of the instrument used in the main study, the following analyses of the grammar test were carried out both at the test and item level.

1) Facility value (FV) or percentage of correct answers that shows the level of difficulty at the item level. According to traditional test theory, item facility values between 20 and 80% are acceptable for norm-referenced tests, but FV requirements may vary according to the purpose of a test (Davies et al., 1999; Henning, 1987).
2) Cronbach's alpha is a test-level statistic that shows the test's reliability if an item is deleted individually from the test. It indicates whether a particular item contributes to the test's internal consistency, with acceptable values > .7 (Taber, 2018).
3) Corrected item-total correlation (CITC) is an item-level statistic that shows item discrimination. It indicates the degree to which an item discriminates between test takers, with acceptable values > .2 (Davies et al., 1999).

Item-level analyses were conducted to determine properties of test items such as item difficulty and item discrimination, to make distinctions between test-takers within a certain range of difficulty, and to determine the test's reliability and validity. Table 4.3 shows item-level analyses of the grammar test, including: 1) facility values, 2) Cronbach's alpha if item deleted, and 3) corrected item-total correlation.

1) Facility values of the 30 test items range from 4.19% to 93.02% (Table 4.3). More specifically, there are 11 items with facility values above 80%, which appears to show that test-takers found them to be fairly easy. However, there are two factors that need to be considered when looking at the facility values of these 11 items: first, the effect of the multiple-choice question test format and, second, the relationship between facility value and Cronbach's alpha if item deleted. The multiple-choice question format affects acceptable facility values, and Davies et al. (1999, p. 96) argue that "for multiple-choice tests the average item difficulty index is set higher" due to the probability that

candidates could guess some of the answers. This may partly explain the somewhat higher facility values of these 11 items. Secondly, further scrutiny of each item regarding its contribution to Cronbach's alpha if deleted sheds more light. Even though the facility values of 11 items exceed 80%, 10 of them contribute positively to the test's internal consistency as shown by Cronbach's alpha if item deleted (Table 4.3). For instance, item 12, with a facility value of 83.24%, contributes positively to the internal consistency of the test because if it is deleted, Cronbach's alpha drops to .743, a value lower than the test-total alpha of .753. In addition, the same item has a corrected-item total correlation of .328 (acceptable >.2), which shows that it discriminates reasonably well among test-takers. Among the 11 items with facility values over 80%, only one (i.e. item 15) does not contribute to the overall alpha of the test, but the difference is 0.005, which can be interpreted as negligible.

Table 4.3 Grammar test: item-total statistics (n = 358)

Variable name	Facility Value (FV)	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
GK_q1	4.19%	.213	.750
GK_q2	28.77%	.468	.734
GK_q3	65.08%	.252	.747
GK_q4	63.41%	.155	.753
GK_q5	16.76%	.415	.739
GK_q6	31.28%	.279	.745
GK_q7	57.54%	.370	.739
GK_q8	70.95%	.016	.761
GK_q9	37.43%	.502	.730
GK_q10	76.26%	.284	.745
GK_q11	93.02%	.303	.746
GK_q12	83.24%	.328	.743
GK_q13	87.71%	.248	.747
GK_q14	83.52%	.185	.750
GK_q15	89.66%	-.008	.758
GK_q16	59.22%	.140	.755
GK_q17	82.40%	.256	.747
GK_q18	90.22%	.231	.748
GK_q19	64.53%	.245	.748
GK_q20	91.06%	.173	.750
GK_q21	66.76%	.373	.739
GK_q22	77.93%	.327	.743
GK_q23	52.79%	.167	.753
GK_q24	89.66%	.293	.746
GK_q25	84.64%	.269	.746
GK_q26	71.23%	.290	.745
GK_q27	77.65%	.382	.739
GK_q28	85.47%	.387	.740
GK_q29	31.56%	.190	.751
GK_q30	30.45%	.252	.747
Cronbach's alpha ***(for all 30 items)***	.753		

On the other hand, the facility values of two items in the grammar test are lower than 20% (4.19% for item 1 and 16.76% for item 5), suggesting that the test-takers found them difficult. For a balanced consideration of the two items it is necessary to check their contribution to Cronbach's alpha if the item is deleted, which reveals that their contribution to reliability of the test is positive, and this appears to counterbalance the lower facility values.

2) A comparison of items according to their Cronbach's alpha if deleted shows that 25 out of 30 items contribute positively to test-overall alpha (Table 4.3).

If any of these items were deleted individually, this would result in a lower test-overall alpha. Items 4, 8, 15, 16 and 23 fail to contribute to test-overall alpha, but it may be reiterated that the difference is quite small (e.g., a test-overall alpha of .753 compared to item 16's alpha if deleted of .755).

3) Corrected item-total correlation is considered, bearing in mind the generally accepted value > .2 (Table 4.3). Eight out of 30 items have lower values, but apart from two items (8 and 15) they are very close to the required limit and some even exert a positive contribution to the test-total alpha. For this reason, they do not appear to be problematic. On the other hand, items 8 and 15 have corrected-item total correlation further below the standard cut-off point (.016 for item 8 and −.008 for item 15) and neither contributes to the test-total alpha. Although a negative CITC is not considered acceptable, the differences account for only 0.008 or 0.005 respectively, which is negligible. In addition, all instruments were tested in the pilot study and proved to be reliable, therefore I decided not to exclude any of the items although their corrected-item total correlation coefficient is below 0.2.

4.6 The background knowledge test

The background test linked to the texts used for the reading comprehension test and the think-aloud procedure was developed in several steps. Subject experts wrote test items to assess the knowledge of those concepts they thought were crucial for the understanding and interpretation of the texts. As with the other tests in the battery, the background knowledge test was pre-trialled in order to produce the final instrument. As discussed in Chapter 2 and Section 4.2, most past studies measured background knowledge indirectly through: a) subjects' study discipline (Alderson & Urquhart, 1983; Koh, 1985; Peretz & Shoham, 1990; Ja'far, 1992; Clapham, 1996; Horiba & Fukaya, 2015), b) subjects' self-assessment of topic or content familiarity (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Lin, 2002; Brantmeier, 2003; Salmani-Nodoushan, 2003; Pulido, 2007; Eidswick, 2010; McNeil, 2011; Lahuerta Martinez, 2013; Shin et al., 2019), and c) a combination of different measures: current, past or future study field, topic familiarity report, reading habits, and reading interest (Bügel & Buunk, 1996; Krekeler 2008; Chigayeva, 2000). More recent studies turned to measuring background knowledge directly by using a background knowledge test (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020; Song & Reynolds, 2022).

However, the texts used for reading comprehension were generally not assessed for specificity. Clapham (1996) noted that using texts that are not specific enough is a limitation of studies dealing with discipline-related knowledge. This may lead to the fact that the effect of background knowledge was not detected. Bearing in mind these two potential limitations, the aim of this study was to measure background knowledge directly through a background knowledge test developed by subject experts and targeted specifically to the knowledge they deemed vital for the comprehension of the texts used in the study.

4.6.1 Aim

The main aim was to develop and validate the background knowledge test based on questions written by subject experts that specifically targeted the knowledge they deemed vital for the comprehension of the texts used in the study. The test was administered in a pre-trial, which resulted in the timing being adjusted and a statistical analysis conducted. The modified instrument was reliable and practical.

4.6.2 Method

The background knowledge test used in the preliminary study contained 15 short-answer questions that were phrased in a way so that they did not raise students' background knowledge and thus confound the variable (Appendix A-15). The questions were matched to the concepts that subject experts deemed to be part of specific knowledge necessary for understanding the texts.

4.6.3 Participants and procedure

A detailed account of the participants and procedure is later given in Sections 4.7.8. and 4.7.9 for reading comprehension testing, and it also applies to background knowledge testing.

4.6.4 Results

Facility values of 15 items were relatively low, ranging from .10 to .75. This was expected, as the test was developed in a way that would show whether students possess specific knowledge of finance that subject experts considered important to comprehend the given texts. Appendix A-16 gives item statistics in the preliminary study. The Cronbach's alpha of all 15 items is .749. Cronbach's alpha of 14 items if question 9 is deleted is .789. If the test is reduced to 12 items by reducing questions 13, 14 and 15, Cronbach's alpha goes down, so this was not taken on. To maintain the highest Cronbach's alpha of .789, question 9 was deleted, and the modified test to be used in the main study included 14 short-answer questions.

4.6.5 Summary of changes to the instrument

The background knowledge test development and pre-trial administration resulted in the following steps and modifications of the instrument for the main study:

1. Development of a 15-item test of specific background knowledge that subject experts considered necessary for understanding the texts used for reading comprehension. The background knowledge test was specifically matched to the reading texts and developed by subject experts (Appendix A-15).
2. Omission of item 9 to maintain the Cronbach's alpha at .789 (Appendix A-16). The final instrument for the main study was reduced to 14 items (Appendix B-2). This reduction improved the reliability and practicality of the instrument.
3. The instructions in the background knowledge test for the main study were amended by adding a sentence giving the students the option of answering questions in either their L1 or L2 (Appendix B-2).
4. As stated before, care was taken that the instructions for administering the test battery for invigilators contained a clear and strict order of test administration (Appendix B-7). This meant that test-takers were to take the background knowledge test prior to the reading comprehension tests.

4.6.6 Coding of answers and calculation of background knowledge scores (main study)

The final 14-item short-answer question test of finance knowledge was taken by all students (n = 358), and their responses were coded using a six-point ordinal scale with the codes and values that would allow for the following purposes:

1) Distinguishing between test-takers with different levels of background knowledge:

   Code 1 = *correct* (3 pts),

   Code 2 = *largely correct* (2 pts),

   Code 3 = *partly correct* (1 pt);

2) Distinguishing between test-takers without background knowledge:

   Code 4 = *incorrect* (0 pts) – wrong background knowledge,

   Code 5 = *don't know* (0 pts) – no background knowledge,

   Code 6 = *irrelevant* (0 pts) – inappropriate background knowledge.

It was decided that such a scale was justified in order to generate information about different levels of knowledge. If the subsequent statistical analyses required it, the codes could be collapsed into new variables. For instance, it was anticipated that the new variable *absence of knowledge* could combine codes 6 = *irrelevant*, 5 = *don't know,* and 4 = *incorrect*.

In coding and validation of students' answers, the question emerged how to consider code 6 = *irrelevant.* To illustrate, code 6 = *irrelevant* covered the answers indicating knowledge that was not appropriate to the context. For instance, a number of students wrote that 'a bond is a connection between two people' which is a perfectly sound general definition, but it fails to capture the correct meaning in financial English. Because the students were explicitly informed that the questions they were asked were related to finance it was concluded that code 6 = *irrelevant* could be collapsed within *absence of knowledge*. Second, codes *correct*, *largely correct,* and *partly correct* as ordinal variable categories indicated different levels of background knowledge. After the consultation with financial experts and statisticians it was concluded that while the individual indicators are of ordinal scale, they can be combined into one construct (an index of individual indicators of background knowledge), which would represent a composite numeric measure of background knowledge.

4.6.7 Reliability of the final measurement instrument (main study)

The reliability of the finance test used in the main study was established by following the same criteria as for the grammar test and reading comprehension test, including:

a) Cronbach's alpha to indicate the test's reliability and internal consistency,

b) Corrected item-total correlation to show item discrimination.

c) Cronbach's alpha if item deleted to show how individual item affects test's reliability and consistency if this item is removed.

Item-level analyses were carried out and the results are shown in Table 4.4.

1) 12 out of 14 items in the finance test contribute positively to the test-total Cronbach's alpha because if any of these items were deleted, this would result in a lower test-total Cronbach's alpha than the alpha for the 14 items (i.e. it would be lower than .716). On the other hand, items 7 and 9 do not contribute to internal consistency. If they were deleted individually, test total alpha would increase by .002 with respect to item 7 and by .004 with respect to item 9. However, these values are so small that they can be considered insignificant.

2) Corrected item-total correlation of 12 out of 14 items is acceptable at values >.2. Again, item 7 and 9 fail to fit the >.2 requirement, with .147 for item 7 and .087 for item 9. In spite of the low CITC, it is noteworthy that the correlation is still positive, and omitting these two items would not significantly improve the discrimination.

Table 4.4 Background knowledge test: item-total statistics (n = 358)

Variable name	**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
BK_q1	.416	.690
BK_q2	.320	.705
BK_q3	.433	.692
BK_q4	.523	.676
BK_q5	.357	.701
BK_q6	.422	.693
BK_q7	.147	.718
BK_q8	.314	.708
BK_q9	.087	.720
BK_q10	.430	.687
BK_q11	.297	.704
BK_q12	.325	.703
BK_q13	.323	.701
BK_q14	.297	.706
Cronbach's alpha ***(for all 14 items)***	**.716**	

4.7 The reading comprehension test

The reading comprehension test was developed in several phases, starting with the selection of reading texts, continuing with item writing and trialling, and wrapping up with analysis and modification of the final instrument.

4.7.1 Aim

The main aim was to develop and validate the reading comprehension test based on four subject-specific texts. To do that the following steps were taken:

1. Selection of texts by subject experts applying the criteria of specificity, familiarity, appropriacy, length, and source-discourse (in two stages) by using a rating instrument with a seven-point Likert scale.
2. Item writing and selection by subject experts (short-answer questions).
3. Pre-trial administration of the instrument and timing of administration.
4. Statistical analysis of pre-trial results and modification of the instrument for the main study to make it reliable and practical.

4.7.2 Method

The specifications for the reading comprehension test used in this study were adapted from Weir's EAP framework for designing reading comprehension tests (Weir, 1993) and are presented in Table 4.5. A detailed description is given in the next section.

Table 4.5 Reading comprehension test specifications

READING OPERATIONS	**Reading carefully at the global level** to understand the main ideas and supporting details; explicit and implicit main ideas	
	Reading carefully at the local level to infer the meanings of lexical items, discourse indicators, and pronominal references	
PERFORMANCE CONDITIONS	**Text-related**	**Text type**: research article or specialist magazine article
		Text length: adequate length that allows for the reading operations
		Text organization: expository text
		Text topic: highly specialist in order to differentiate among readers with varying degrees of relevant background knowledge
	Test-related	**Test purpose**: to assess reading comprehension of authentic subject-specific texts used in the academic context
		Test structure: according to the sequence in which the relevant information appears in the text
		Test items: short-answer questions, as the most common format for testing comprehension at global and local levels with a high degree of ecological validity
		Timing: pre-trialled

4.7.3 Text selection

The selection of texts for reading comprehension was carried out in the following three stages:

1. The main subject expert, a professor of business finance, was asked to select about ten research and specialist magazine articles according to the following criteria:

a) High specificity of the texts;
b) Appropriacy for the target population;
c) Presence of new knowledge that the students are not expected to have;
d) Absence of tables or graphs;
e) Possibility to use a text from the article or the complete text.

The choice of research articles and magazine articles was motivated by the fact that these two formats make up students' coursework reading, so it was assumed that they ought to be familiar with the genre. The rationale for the choice of texts was to maintain ecological validity, including the authenticity of the text and task. The main subject expert selected passages from six research articles taken from the *Journal of Finance*, the *Journal of Financial Economics*, the *Financial Management*, the *RAND Journal of Economics*, and five articles from the newspaper *The Economist* (Appendix A-7).

2. In the next phase, a rating instrument using a seven-point Likert scale was designed (Appendix A-8). The wording of questions was validated by a statistician specializing in questionnaire design. The instrument contained five questions asking expert raters to assess each of the 11 texts for five criteria: 1. specificity, 2. appropriacy, 3. familiarity, 4. new knowledge, and 5. explanation of the new knowledge.

3. After three finance experts assessed all 11 texts, their ratings were tabulated (Appendix A-9). Then the means were calculated for each text as summarized in Table 4.5. The aim was to select four articles for reading comprehension item writing that would be followed by a pre-trial on the target population. Although it was clear that because of time constraints it was likely that eventually only three texts would be used, it was felt that it would be better to select four texts in the first round, and then eliminate the text with the lowest reliability after the pre-trial on the target population.

Table 4.6 Text selection: the calculated means and standard deviations of five criteria used for expert assessments of 11 texts

	Specificity		Appropriacy		Familiarity		New Knowledge		Explanation	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
OPEN MARKET STOCK R.	**6.67**	**0.47**	**3.33**	**1.89**	**3.33**	**1.25**	**6.00**	**0.00**	**3.67**	**1.70**
STAR-STRUCK	**6.33**	**0.47**	**5.33**	**0.94**	**3.00**	**0.82**	**5.67**	**0.47**	**4.67**	**0.94**
BONDS THAT ROCK & R.	**6.00**	**0.82**	**4.33**	**1.25**	**4.00**	**0.82**	**5.67**	**0.94**	**5.33**	**1.25**
TAKING STOCK	**6.00**	**0.82**	**4.67**	**1.25**	**4.00**	**1.63**	**4.33**	**1.70**	**4.67**	**1.25**
FINANCIAL DECISIONS	5.33	0.47	4.67	1.25	5.33	0.47	5.00	0.82	4.33	1.25
THE TIMING OF IPO’S	4.67	1.25	5.00	0.82	5.00	0.82	5.00	0.00	5.00	1.41
AGENCY COSTS	4.67	1.25	6.00	1.41	5.33	1.70	4.33	1.25	4.00	1.41
FLOTATION FEVER	4.67	0.47	5.67	0.94	5.33	1.70	3.33	1.25	4.33	1.25
GREENMAIL	4.67	1.25	5.00	1.41	5.00	0.82	3.33	1.25	5.33	2.36
FLOATING IN THE AIR	3.67	0.94	6.00	0.00	4.33	0.47	4.67	0.47	5.33	0.47
CHAMELEON BONDS	4.00	1.41	6.33	0.47	5.67	1.25	3.33	0.94	3.67	1.70

Note: *M* = Mean; *SD* = Standard Deviation

Table 4.6 shows the expert assessment of 11 texts according to the five criteria, starting with specificity. As the specificity of articles critically influences the effect of background knowledge on reading comprehension (Clapham 1996), it was decided that “specificity” would be the primary selection criterion of the articles to be used in the pre-trial. Four articles had means of 6.0 to 6.66 (on a scale of 1–7), thereby showing that the raters found them to be highly specific (“Star Struck” (SS), “Bonds that rock and roll” (BRR), “Taking stock” (TS), and “Open market stock repurchase signalling” (OMS)). This was also corroborated by raters’ assessment of familiarity. The familiarity ratings for the same articles were between three and four, which is lower than the ratings for the other articles and indicates that subject experts expect these four articles not only to be highly specific, but also to be least familiar to the students. A similar trend was evident also from the “New Knowledge” ratings, which are higher than for the remaining articles, with 5.66 to 6.0 for SS, BRR, and OMS, whereas only TS had a somewhat lower rating of 4.33. The “explanation of new knowledge” ratings did not seem to indicate any particular trend, but the least explanation is given for CB (3.66) and the rest of the articles ranged from 4.0 to 5.33. Similarly, the appropriacy ratings did not seem to reveal a specific trend; it is noteworthy that OMS had the lowest rating of 3.33. Still, this rating is approximately mid-scale and thus the article could not be ruled out completely. The comparison of expert ratings by individual raters is summarized in Appendices A-9, A-10, and A-11. Short interviews with experts were conducted, resulting in the selection of the following four articles for the pre-trial: “Star-Struck” (SS), “Bonds that Rock and Roll” (BRR), “Taking Stock” (TS), and “Open Market Stock Repurchasing” (OMS).

4.7.4 Readability indices of texts

The four texts to be used in the preliminary trial were tested for their readability as shown in Table 4.7. To estimate the texts’ lexical load, two readability indices were calculated: the Flesch Reading Ease and the Gunning Fog index. Originally developed for L1 reading, they are based on the total numbers of words, complex words, sentences, and syllables in the text. The Flesch Reading Ease index measures word length, while the Gunning Fog index measures the number of syllables per 100 words and per sentence. The underlying assumption is that long words and sentences are more difficult than short ones. However, it must be noted that the difficulty of a text also results from aspects other than the ones these calculations are based on. For this reason, readability indices can be taken only as a very rough measure of difficulty (Alderson, 2000, p. 73) and they should be paired with other assessments of difficulty that take into account the rhetorical features of a text, assumed background knowledge or level of specificity, conceptual density, amount of embedding, or amount of implicit or explicit ideas that make the text either opaque or transparent.

As a preliminary indication of text difficulty, the calculations for the four texts were checked on the Flesch Reading Ease rating scale, whereby scores of 90 to 100 indicate a text that is very easy, 80 to 89 easy, 70 to 79 fairly easy, 60 to 69 standard, 50 to 59 fairly difficult, 30 to 49 difficult, and 0 to 29 very difficult. According to this scale, OMS and TS (25, 23) are ranked as “very difficult” texts, whereas SS and BRR (43, 40) are “fairly difficult”. These findings were not surprising and were actually also reflected in the sources of the articles: the texts rated as “very difficult” (TS and OMS) were research articles from scientific journals, whereas the texts rated as “fairly difficult” (SS and BRR) were taken from a specialist magazine. Another check was made by calculating the Gunning Fog index, whereby texts with scores above 12 are considered “very difficult” to read. Academic papers would be around 15 to 20 and newspaper articles around 14. Again, the Gunning Fog indices for the battery of four texts reveal similar results to the Flesch: TS, OMS, and BRR (18, 16, 14) are “very difficult" and SS is just below this point (11). As these readability indices were originally indicators of L1 reading, it must be acknowledged that the difficulty level experienced by L2 readers may cause even more difficulties.

Table 4.7 Readability analysis

	Star-Struck	Bonds that Rock and Roll	Taking Stock	Open Market Stock Repurchase Signalling
Total words	425	1,282	581	1,025
Complex words	54	192	129	184
Total sentences	26	57	25	42
Total syllables	735	2,173	1 097	1,896
Flesh Reading Ease	**43.94**	**40.61**	**23.51**	**25.57**
Gunning Fog index	**11.62**	**14.99**	**18.18**	**16.94**

4.7.5 Reading comprehension test development

After the selection of the four texts, two subject experts were asked to write items for the RC test – about five to six items for each of the four texts – so that the total test would consist of about 20 to 24 items. The agreed format was short-answer questions that would check readers' global and local understanding of main ideas and supporting details. The items that subject experts wrote were discussed at a meeting and the selection of items was made with mutual agreement.

Item types

To assess reading comprehension, several item response types can be used with varying degrees of reliability and validity. Of the 20 standardized reading assessment tasks that Grabe (2009, p. 359) lists, the most frequently used are short-answer questions (SAQs), multiple-choice questions (MCQs), dichotomous items (True/False), cloze, free recall, and summarizing. The intention in this study was to minimize the "method effect", so the pros and cons of item types were considered before the final selection of the item type. All item response types to test reading comprehension have some disadvantages. To list just the most important, SAQs require the test-takers to engage in the productive activity of writing, which may confound the measure of reading comprehension. A possible way to reduce this effect is by limiting the length of answers. On the other hand, MCQs do not require any writing on the part of the test-taker, so they are fairly objective and easy to score. However, a big disadvantage of MCQs is that they risk being able to be answered without reading, and, depending on the number of given choices, there is a good possibility of selecting the correct answers by chance. The main downside of dichotomous items is a 50% chance of getting the item right without actually showing any reading comprehension ability. Although the chance factor can be reduced by including the "Not given" option, the odds of getting items right are so high that this was not the best item type to assess reading comprehension for the purpose of this study. Another popular method, cloze, has been criticized in relation to construct validity. It is not clear what skill it actually measures, whether it can measure beyond local comprehension information, and whether it only measures reading comprehension (Trace, 2020).

Surveying the disadvantages of most commonly used item types pointed out that the best choice for the present study was SAQs. Despite the fact that they require the reader to write, SAQs can be limited to the shortest possible answers. In the first place, they have a high level of ecological validity as regards reflecting a real-life activity and this is their major strength. Yet, the potential variety of answers may affect the marking of such items and this should be considered in item writing and scoring the correct answers. Another upside of SAQs is the fact that they may require the reader to show comprehension at either the word, sentence, or text levels, or make connections between the information on both local and global levels. Finally, to minimize the language effect, it is possible to give the test-takers the choice of writing their answers in L2 or L1, which is what was done in this study.

4.7.6 Preliminary trial

As presented in Appendix A-13, the reading comprehension test of four texts was administered to the target population together with the remaining parts of the test battery (i.e. grammar knowledge test, background knowledge test, and post-reading questionnaire). The trial had two main aims: the first was to check the reading comprehension test for difficulty and reliability through the statistical analysis of the results. The second aim was to determine adequate timing for reading comprehension and other tests by measuring the time that students take to complete the tasks. Table 4.8 lists the range of timings for each section and the subsequent plan that was made for the main study.

Table 4.8 Timing and modifications of the test battery: Preliminary and main studies

	Grammar test	Background knowledge test	Reading comp. test + Post-reading questionnaire **SS**	Reading comp. test + Post-reading questionnaire **BRR**	Reading comp. test + Post-reading questionnaire **TS**	Reading comp. test + Post-reading questionnaire **OMS**
Measured timing **Preliminary study** ca 110 min	15–20	15–20	10	15–20	10–20	10–20
Adjusted timing **Main study** 80 min	10 GK test reduced from 60 to 30 items.	20 BK test reduced from 15 to 14 items.	10	20	20	0 OMS text eliminated, consequently RC test battery reduced from 22 to 17 items.

4.7.7 Statistical analysis

Appendix A-14 summarizes the results of the preliminary reading comprehension test of 20 students. Students' answers were scored on a 6-point scale: 1 correct, 2 largely correct, 3 partly correct, 4 incorrect, 5 not given, and 6 irrelevant. This scale was adopted because it was felt that there may be different degrees of correctness of the given answers. The scale also captures the fact that some students did not answer some questions, or they gave irrelevant answers. It was felt that, if necessary, points 4 incorrect, 5 not given, and 6 irrelevant could all be collapsed into point 4 incorrect. The results of the statistical analysis of 22 items for four texts showed the following trends:

1. Four items had facility values out of the acceptable range of 20-80%. Specifically, three items had 100% FV and one item had 5 %, which indicated that these items were either too easy or too difficult, so it was best to take them out of the test battery.
2. Cronbach's alpha was used as an indicator of the test's reliability, and as shown in Appendix A-14, it was 0.516 for all 22 items. This is a low but still acceptable value of Cronbach's alpha. However, because the total test battery took 110 minutes to administer, as shown in Table 4.8, it was necessary to decide where to make cuts to the test. Two solutions were at hand. The first one was omitting one reading comprehension text and its corresponding test items, thus cutting down the administration by 20 minutes. Cronbach's alpha if five items were deleted would be 0.536, which is slightly higher than the Cronbach's alpha for the total test comprising 22 items (.516). The second option was omitting items with 100% facility plus the items related to the last text. This would render a Cronbach's alpha of 0.517, which is almost the same as Cronbach's alpha for the total test of 22 items (.516).

To shorten the administration while maintaining the highest possible Cronbach's alpha, it was decided to eliminate the OMS text and thus reduce the reading comprehension test from 22 to 17 items, with a Cronbach's alpha of 0.536. Two items were kept despite their 100% facility because they contributed to a higher internal consistency. If they had been deleted, the test alpha would have dropped to 0.498.

4.7.8 Subject experts and participants

Subject experts were involved at all stages of the text selection and test development. One expert provided a selection of 11 texts according to the criteria of specificity, length, appropriacy, and new knowledge as described in Section 4.7.3. Three experts rated the texts according to the selected criteria as explained before. Two experts wrote test items and selected which ones should be included in the preliminary trial. They were also involved in the assessment and validation of students' answers that was the basis for the subsequent statistical analysis.

The sample consisted of 20 students (six males and 14 females) that took part in Preliminary Study 2. They were administered the battery of four instruments, including the reading comprehension test with 22 items based on four texts. In a separate think-aloud preliminary study, three students trialled the think-aloud procedure while reading the four texts and taking the reading comprehension test.

4.7.9 Procedure

As outlined in Section 4.4, the reading comprehension test was part of the test battery that was administered to students outside of their course time. The grammar test and the background knowledge test were administered first, followed by reading comprehension test and the post-reading questionnaire. It was crucial to keep this sequencing of administration, and the following two conditions in particular. First, students were given the background knowledge test prior to the reading comprehension test in order not to confound the two variables. Second, students were given each of the four reading comprehension texts separately and were invited to read them, answer the reading comprehension questions, and fill in the corresponding post-reading questionnaire ratings for the text. After this was done for one text, the sheets with the text were removed. Then the next text was distributed to the students, and the procedure was repeated with the next text. This sequencing was adopted to ensure that students were not reverting back to the texts they had read before. In addition, it was considered important that test-takers should enter their ratings of topic familiarity, text difficulty, etc., in the post-reading questionnaire immediately after reading a particular text while their perception and memory of a text was still fresh. If they had done this for all four texts together, this could have easily resulted in mixing up the ratings and texts.

Like the grammar test administration, the timing of the reading comprehension test was not fixed, and the students were allowed as long as they needed to read and answer the reading comprehension items comfortably. They were informed that the pre-trial administration was also intended to time the administration of the test battery. There were no considerable differences in the timings of individual students, so it was possible to determine the timing appropriate for administrating the main study.

4.7.10 Summary of changes to the reading comprehension test battery

The reading comprehension test development and pre-trial administration resulted in the following steps and modifications of the instrument for the main study:

1. Shortlisting four texts by using a 7-point Likert scale rating instrument, based on five criteria: specificity, appropriacy, familiarity, new knowledge, and explanation of new knowledge (Table 4.6).
2. Item writing and item selection by subject experts, resulting in a 22-item reading comprehension test based on four texts that was pre-trialled in a test battery together with the grammar test, background knowledge test, and post-reading questionnaire involving 20 students (Appendix A-13).
3. Removing one text to reduce the length of administration. The OMS text was omitted, so that the number of items was cut down from 22 to 17, and the time of administration could be shortened by a further 20 minutes in addition to the cut of 10 minutes due to cutting the grammar test length. As a result, a total of 50 minutes would be allocated for the entire reading comprehension test based on three texts. Because it was pre-trialled on the target population, it was expected that that the time allowed would fit with the reading speed for ESL students of about 200 wpm (Nuttal, 1996).
4. Reordering the sequence of items so that all items followed the order of occurrence of information in the text.
5. Preparing the reading comprehension test with acceptable reliability as shown by the Cronbach's alpha of 0.536.

The refinement of the final instrument to be used in the main study aimed at maintaining its reliability and practicality. In addition, I also prepared the instructions for the invigilators that would administer the test battery in the main trial, so that it contained proper timing and all relevant information in Slovene (Appendix B-7). I also prepared the timetable of administration of the test battery, so that the order of the reading comprehension tests varied in different groups. The aim of this variation was to minimize the "order effect".

4.7.11 Coding of answers and calculation of reading comprehension scores (preliminary study)

The 17-item reading comprehension test was taken by all students (n = 358), and their responses were coded on the six-point ordinal scale that was also used for the background knowledge test and contained the following codes and values:

Code 1 = *correct* (3 pts),
Code 2 = *largely correct* (2 pts),
Code 3 = *partly correct* (1pt),
Code 4 = *incorrect* (0 pts),
Code 5 = *don't know* (0 pts),
Code 6 = *irrelevant* (0 pts).

Similar to the background knowledge test, it was intended that the scale used would capture the different levels of comprehension and make it possible to distinguish between test-takers with different levels of reading comprehension. It would allow the coding of scores necessary for the subsequent linear regression analysis as variables can be treated as numeric.

4.7.12 Reliability of the final measurement instrument (preliminary study)

Similar to grammar knowledge and background knowledge tests, the reliability of the reading comprehension test was checked by looking at three factors:

a) Cronbach's alpha indicating test's reliability and internal consistency,
b) Corrected item-total correlation showing item discrimination.
c) Cronbach's alpha if item deleted showing how individual item affects test's reliability and consistency if this item is removed.

Item-level analyses were carried out and the results are shown in Table 4.9.

1) 16 out of 17 items in the reading comprehension test contribute positively to the test reliability because deleting any of these items would result in a lower test-total Cronbach's alpha than for the test of 17 items (i.e. it would be lower than .695). Only item 1 does not contribute to internal consistency because if it was deleted, the test-total Cronbach's alpha would increase to .700. Yet the

increase of .005 is so small that it can be stipulated that it hardly affects test-total reliability.

2) The corrected item-total correlation (CITC) of 14 items out of 17 is acceptable at values >.2. Three items (items 1, 2, and 12) have CITC values slightly lower than .2, which indicates that they do not contribute to discrimination. However, despite their lower CITC, two items (item 2 and 12) contribute to test reliability as evidenced by Cronbach's alpha if item deleted. Conversely, if item 1 was deleted, Cronbach's alpha would be slightly higher than the test-total alpha, but this difference is very small (.7 compared to .695). Cronbach's alpha for the test of 17 items, including item 1, equals 0.695, which indicates an acceptable degree of test's reliability.

Table 4.9 Reading comprehension test: item-total statistics (n = 358)

Variable name	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
RC_q1	.170	.700
RC_q2	.165	.694
RC_q3	.286	.682
RC_q4	.249	.686
RC_q5	.400	.672
RC_q6	.460	.657
RC_q7	.321	.681
RC_q8	.275	.683
RC_q9	.402	.669
RC_q10	.425	.664
RC_q11	.325	.678
RC_q12	.150	.694
RC_q13	.217	.689
RC_q14	.243	.687
RC_q15	.341	.676
RC_q16	.222	.688
RC_q17	.230	.691
Cronbach's alpha ***(for all 17 items)***	**.695**	

4.8 The post-reading questionnaire (PRQ)

Immediately after reading each of the four texts and completing the comprehension check, test-takers were asked to use a 7-point scale to rate their familiarity with the text topic, how much this topic familiarity helped them to answer the comprehension questions, and how they rated the difficulty of the texts (Appendix A-17). As shown in table 4.9, test-takers were also asked to provide comments to their assessments.

Table 4.10 Post-reading questionnaire: preliminary study

STAR-STRUCK	
How familiar were you with the topic before you read the text?	Not at all 1 2 3 4 5 6 7 A very great deal
If you were familiar with the topic of the text, how much did this help you answer the questions?	Not at all 1 2 3 4 5 6 7 A very great deal
How easy / difficult was the text?	Very easy 1 2 3 4 5 6 7 Very difficult

Your comment

____

____

All test-takers in the preliminary study filled in the post-reading questionnaire and the results are shown in Appendix A-18. Table 4.11 presents the means of their assessments. Test-takers found the BRR text the least familiar and also the most difficult. On the other hand, they rated the TS text as the most familiar, reporting that this helped them considerably. Test-takers' ratings were paired up with the statistical analysis of reading comprehension tests and expert opinion, and this resulted in scrapping the OMS text. The decision was to keep the most contrasting texts in terms of familiarity and its effect.

It was planned that in the main study test-takers' familiarity and other ratings would be compared to their scores on tests of background knowledge and reading comprehension, and in this way explore how self-reported perceptions can be matched to test-takers' actual performance.

Table 4.11 Post-reading questionnaire, preliminary study: mean assessments

Text	Familiarity	Effect of familiarity	Text difficulty
SS	2.2	3.5	3.6
BRR	2.0	3.3	5.2
TS	3.2	3.9	4.3
OMS	2.6	3.5	4.9

Having discussed the PRQ questions and results with subject experts, the PRQ for the main study was modified (see Appendix B-5). At the beginning, two baseline questions were added about the test-takers' general and course-related reading habits. A sentence was added stating that test-takers could provide answers either in L1 or L2. The questions about topic familiarity and text difficulty were reordered, and a question about students' assessment of how interesting they find the text was added as shown in Table 4.12. The open question asking test-takers to comment was moved after the assessment for each text. In addition, the question to comment was rephrased to be more specific. Instead of just the instruction: *Comment*, test-takers were asked specifically: *What do you attribute the level of difficulty to?* and: *In what way did your familiarity with the topic or lack of it show in your reading comprehension?* It was felt that more specific questions would elicit more detailed answers from the test-takers that would shed better light on the studied issues. The differences in PRQs in the preliminary and main studies are shown in Appendices A-17 and B-5.

Table 4.12 Post-reading questionnaire: main study

STAR-STRUCK

How **difficult** did you find the text?	Very easy Very difficult **1 2 3 4 5 6 7**

Why? What do you attribute the level of difficulty to? Please comment.

____

How **familiar** were you with the topic before you read the text?	Not at all A great deal **1 2 3 4 5 6 7**
How much did this **affect** your reading comprehension?	Not at all A great deal **1 2 3 4 5 6 7**

In what way did your familiarity with the topic or lack of it show in your reading comprehension?

Please comment. ____

____

How **interesting** did you find the text?	Not at all A great deal **1 2 3 4 5 6 7**

4.8.1 Reliability of the measurement instrument for Self-reported Familiarity, Difficulty, and Interest

In this study, I aimed to compare a) readers' background knowledge assessed with a finance test, and b) readers' self-reported familiarity with text topics in terms of their effect on reading comprehension (see Section 4.1 for the literature review). To do that, a composite index for Self-reported Familiarity was created by combining the indicators of familiarity with: 1) ”Star-Struck” topic, 2) ”Bonds that Rock and Roll” topic, and 3) ”Taking Stock” topic.

I created composite indices also for “Difficulty All Texts” and “Interest All Texts” using the same approach. Composite indices were created as follows:

1. Self-reported Familiarity construct (all texts): combining SS familiarity, BRR familiarity, and TS familiarity into one composite measure.
2. Difficulty construct (all texts): combining SS difficulty, BRR difficulty, and TS difficulty into one composite measure.
3. Interest construct (all texts): combining SS interest, BRR interest, and TS interest into one composite measure.

Self-reported Familiarity construct, which is a measure of average familiarity with the three text topics, will be used as an independent variable in regression models addressing the research questions in Chapter 5. The Difficulty and Interest as constructs will be used in Section 5.9 as part of Post-reading Questionnaire analysis.

As with the grammar knowledge, background knowledge, and reading comprehension tests, the reliability of Self-reported Familiarity (SRF) measure, Difficulty measure, and Interest measure was checked by looking at: a) Cronbach's alpha to indicate the test's internal consistency, and b) corrected item-total correlation to show item discrimination and internal consistency. Each of the three items for each concept (SRF, Difficulty, Interest) was analysed for those two parameters. Response rate for each concept was as follows: topic familiarity (n=354), difficulty (n=353), and interest (n=313). Cronbach Item-level statistics for the reading comprehension test are presented in Table 4.13. The total Cronbach's alphas were 0.525 (SRF), 0.552 (Difficulty) and 0.525 (Interest) – it can be assumed that the instruments are reliable and the new constructs can be used in subsequent statistical analyses. Also, all corrected item-total correlation coefficients for all three concepts from Table 4.13 were > .2.

Table 4.13 Self-reported Familiarity, Difficulty, Interest: item-total statistics

Text	**Self-Reported Familiarity**		**Difficulty**		**Interest**	
	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
SS	0.417	0.286	0.433	0.332	0.337	0.425
BRR	0.338	0.427	0.346	0.477	0.419	0.279
TS	0.269	0.541	0.313	0.525	0.263	0.538
Cronbach's alpha *(for all 3 items)*	0.525		0.552		0.525	

4.9 The think-aloud procedure (preliminary study)

The think-aloud procedure was intended to provide qualitative data that would add insights into the reading comprehension process. These were to be combined with the quantitative data obtained from the testing study in order to offer answers to the research questions.

4.9.1 Aim and method

The aim of the think-aloud preliminary procedure was to acquaint test-takers with the think-aloud procedure, offer them a brief training in how to think aloud, and then conduct the think-aloud procedure with the four texts selected.

4.9.2 Participants and procedure

Two students (one male and one female) took part in the preliminary think-aloud study. The think-aloud procedure was administered in a single session with each student, starting with a brief training session in thinking aloud and instructions (Appendix

A-19). The text used for training comprised two paragraphs and was taken from a finance magazine selected by the main subject expert. This training was followed by the think-aloud procedure for the four texts. I anticipated that the test-takers would be somewhat inhibited in thinking aloud and would need some encouragement and support to think aloud.

4.9.3 Results

The procedure showed that the selected texts were suitable for thinking aloud. The think-aloud data provided evidence to draft an exploratory coding scheme that would inform the design of the coding scheme in the main study. During the think-aloud procedure, the students needed some reminding to think aloud. The trial showed that the battery of four texts was too long, because both test-takers reported being tired and they became much slower when reading the last text. It was clear that fatigue could influence the results, so this was a contributing factor in the decision to eliminate one text from the battery and thus cut the administration time.

4.10 Research governance and ethics

To assess the ethical soundness of this research, I used the guidelines from the Institute of Education's Code of Practice for Research Degrees as well as the BERA revised ethical guidelines for educational research (http://www.bera.ac.uk/). The ethics process was approved by the School of Business and Economics University of Ljubljana and written approval was sought and given by the Vice-Dean of the School of Business and Economics University of Ljubljana. The process was also in line with the regulations at the Institute of Education at the time of research. I applied the principles underpinning the guidelines to demonstrate respect for the person, knowledge, democratic values, the quality of educational research, and academic freedom. As described in the next sections, I self-assessed the research regarding my responsibilities to three main groups of stakeholders: participants, sponsors, and the educational research community.

4.10.1 Responsibilities to participants

The participants completed the tests and answered the questions in post-reading questionnaire. To demonstrate my responsibilities to the participants, I followed the principles concerning voluntary informed consent, privacy, the right to withdraw, absence of incentives, and diffusion of tension, as explained next.

a) Voluntary informed consent: This principle was fully taken into account. Students were provided with key information about the study. They were provided with oral explanations about the type, purpose, and procedure of the research, and what exactly their involvement would be. They were asked to participate of their own free will and give oral consent.
b) The right to withdraw: Apart from having the research explained to them, the participants were also told that they could withdraw from the research if they wished to do so. There were two students who decided not to take part in the testing study. Their decision was respected and no effort was made to persuade them to change their minds. In the think-aloud study two students withdrew mid-test and did not complete all sections of the test battery; one person reported being tired and the other had to leave. I did not try to persuade them to carry on because I felt it was important to respect their decisions.
c) Privacy: All participants in the research were told that the data obtained from the research would be treated as confidential and anonymous. This was guaranteed not only verbally but through the procedure itself: no student names were required, and students were given codes on stickers to mark each test they took. This made it possible to identify the tests and data belonging to one individual while still maintaining complete anonymity. On the other hand, students who took the think-aloud study were identified by name, but were promised that in any subsequent reports their names would be treated as confidential and would not be disclosed. The participants were also told that if they wanted to see the research results, they could obtain them directly from the researcher at the end of analysis.
d) Incentives**:** No incentives were used to encourage students to participate.
e) Diffusion of tension: To help diffuse tension and unease during the think-aloud study, test-takers were offered soft drinks.

4.10.2 Responsibilities to sponsors

My research was partly sponsored by the School of Economics and Business, University of Ljubljana, Slovenia, in two ways: by enabling access to participants, and by partly funding my study. I obtained the written consent of the Vice-Dean permitting me to use the student data in order to carry out research at school. In addition, it was agreed that if the sponsor so desired, the results of the study would be provided.

4.10.3 Responsibilities to the educational research community

I made sure that this research met the highest standards by not engaging in any form of misconduct or malpractice as listed in the BERA guidelines for undertaking research in education. Written approval was given by the vice-dean of the faculty that the research could be undertaken, and access to student data was granted.

4.11 Chapter summary and conclusion

This chapter provided a description of the research design and data collection in this study, both in the preliminary and main phases. It discussed how the key research variables were operationalized, and offered a detailed account of how the main instruments were developed, trialled, and modified. These were explored with respect to the participants, materials, procedure, and analyses used. The main aim was to develop reliable, valid and practical instruments that would render qualitative and quantitative results that would answer the research questions about the contribution of background knowledge to the reading comprehension of subject-specific texts and the patterns that readers use in the process.

CHAPTER 5
RESULTS: THE TESTING STUDY

5.1 Chapter aim and overview

This chapter presents the analyses and results of the testing study involving a total of 358 students from the School of Economics and Business, University of Ljubljana. It explores the relationships among three independent variables – Background Knowledge (BK), Grammar Knowledge (GK), and Self-reported Familiarity (SRF) – and one dependent variable – Reading Comprehension (RC) in order to provide answers to the research questions. As presented in Chapter 4 grammar knowledge was operationalized through a score on a grammar test, background knowledge through a score on a finance test, reading comprehension through a score on a reading comprehension test of three texts, and self-reported familiarity through readers' assessment in a post-reading questionnaire. Chapter 5 begins with a presentation of descriptive statistics related to the three main variables in the main study. It looks at the results of bivariate correlation and multivariate regression analyses. The data are examined at the levels of the test, section, and item. The answers to the research questions are proposed and research hypotheses are addressed.

5.2 Grammar test

5.2.1 Introduction

As described in Chapter 4, in preliminary trials the grammar test was reduced from 60 items to 30 in a three-step procedure with the intent of increasing the test's reliability, validity, and practicality. The first step was a construct validation procedure to filter out lexical items and improve construct validity. The second step was a statistical analysis to screen out those items that failed to contribute positively to the overall-test Cronbach's alpha, and had facility values outside the range of 20% to 80%. This allowed only items that discriminate properly between test-takers. Finally, the length of the test and time of test administration was reduced, thereby improving its practicality. The reliability analysis results are provided in Chapter 4.

5.2.2 Aim

The grammar test used in the main study consisted of 30 multiple-choice questions and was taken by 358 students. The main aim of the test was to obtain information about students' grammar knowledge, which would later be used as a predictor of reading comprehension. It also served as a basis for the division of groups (i.e. students with either high, medium, or low grammar knowledge) and explore how they could be compared to groups with high, medium or low background knowledge regarding their performance on the reading comprehension test.

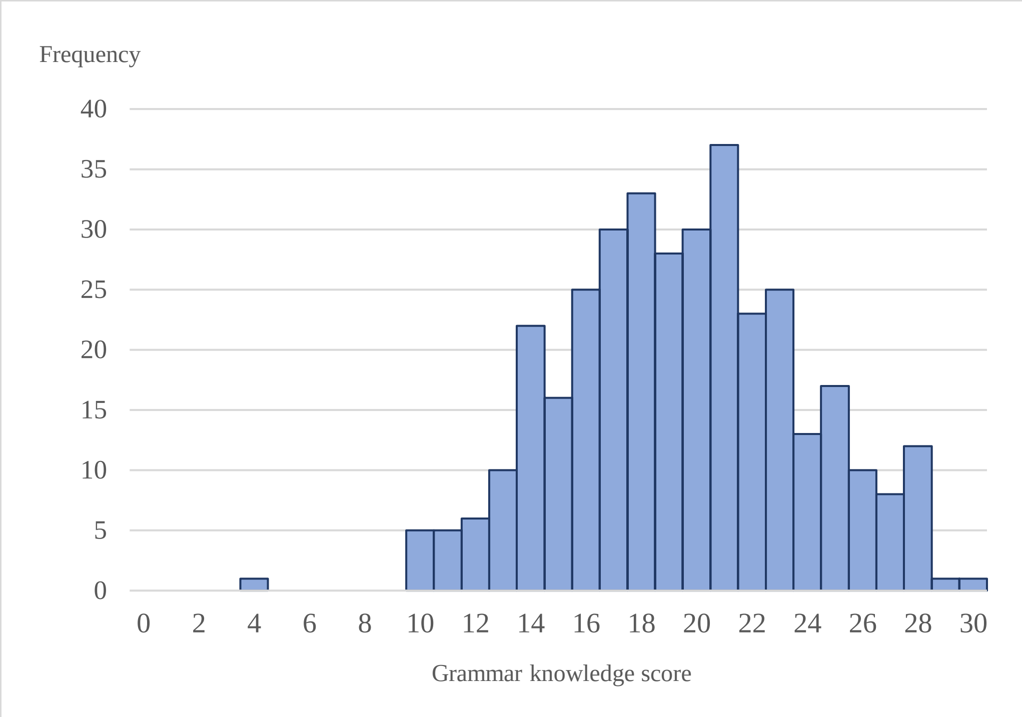
5.2.3 Descriptive statistics

The variable expressing Grammar Knowledge, operationalized through a 30-item multiple-choice grammar test, has a normal distribution (Table 5.1, Figure 5.1). The mean score of all subjects (n = 358) at the test level was 19.44 out of a total score of 30, with a median of 19 and a mode 21. The standard deviation is 4.31. The kurtosis coefficient of −0.182 (with standard error of 0.257) and a skewness coefficient of −0.033 (with standard error of 0.129) can be used to confirm the normality of the distribution. The lowest-scoring tenth of students achieved 14 or fewer points, whereas the top-scoring tenth achieved 25 points or more.

Table 5.1 Grammar test: descriptive statistics

Statistic		Grammar Knowledge
n		358
Mean		19.44
Mean (%)		64.80%
Median		19.00
Mode		21.00
Std. Deviation		4.311
Skewness		-.033
Std. Error of Skewness		.129
Kurtosis		-.182
Std. Error of Kurtosis		.257
Minimum		4
Maximum		30
Percentiles *(selected deciles and quartiles)*	10	14.00
	25	16.00
	75	22.00
	90	25.00

Figure 5.1 Grammar test: frequency distribution



5.3 Background knowledge test

5.3.1 Introduction

As described in Chapter 4, a finance test consisting of 14 short-answer questions was developed in conjunction with four financial experts according to test specifications and results of the preliminary trial. The test taps into students' knowledge of various financial concepts and aspects of finance deemed by these experts to be of crucial importance for in-depth reading comprehension of the selected texts. Care was taken not to confound the variables by including items that would overlap with the items from the reading comprehension test. The reliability analysis results are provided in Chapter 4.

5.3.2 Aim

The aim of the 14-item finance test used in the main study and taken by 358 students was to obtain an indication of students' knowledge of finance/background knowledge, which would be later used as a predictor of reading comprehension. It would also make it possible to distinguish among students with different levels of knowledge of finance and serve as a basis for the division of groups (students with either high, medium, or low background knowledge) and allow comparison with student groups with high, medium, or low grammar knowledge regarding their performance on the reading comprehension test.

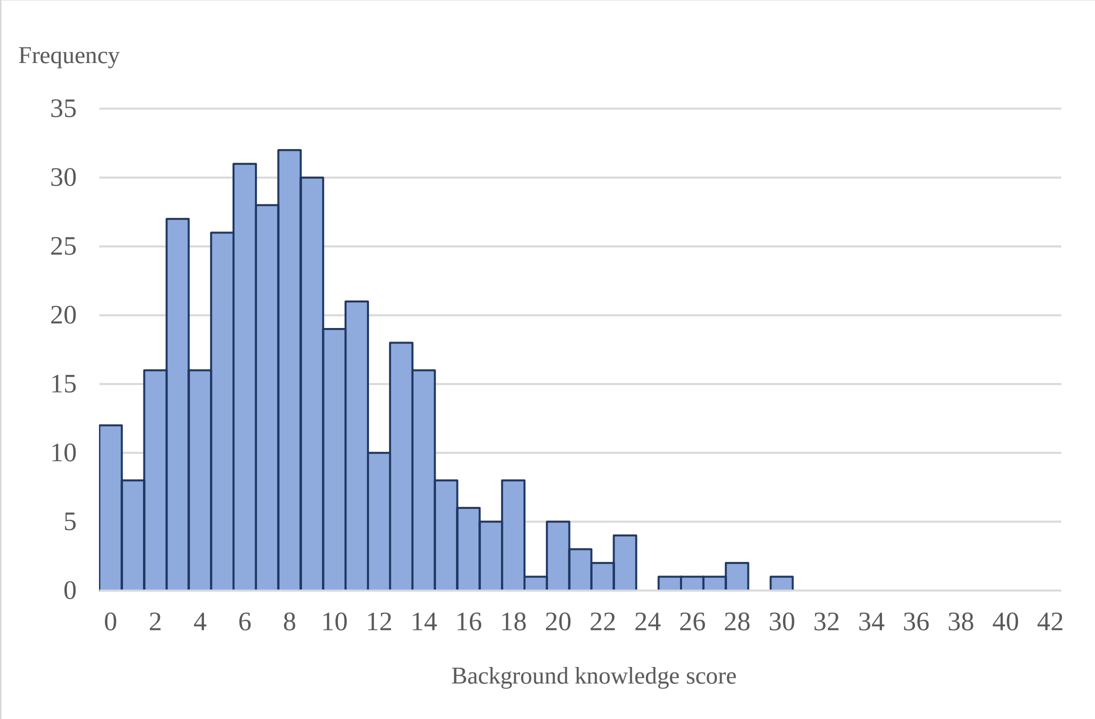
5.3.3 Descriptive statistics

Table 5.2 and Figure 5.2 show the descriptive statistics for the composite index Background Knowledge that was operationalized through a short-answer question finance test with a potential maximum score of 42 points. 14 items were scored by using an ordinal scale with the following coding: 0 points (incorrect, irrelevant, don't know), 1 point (partly correct), 2 points (largely correct), and 3 points (correct). The derived variable is leptokurtic and right-skewed, which is shown by the skewness coefficient of .959 (with standard error of 0.129), and kurtosis coefficient of 1.122 (with standard error of 0.257). The values of the variable (i.e. test scores) range from 0 to 30 out of 42, the mean is 8.80 and standard deviation is 5.598. The median and mode are both 8. The lowest scoring tenth of students achieved 2 or fewer points, whereas the top scoring tenth achieved 16 points or more.

Table 5.2 Background knowledge test: descriptive statistics

Statistic		**Background Knowledge**
n		358
Mean		8.80
Mean (%)		20.95%
Median		8.00
Mode		8.00
Std. Deviation		5.598
Skewness		.959
Std. Error of Skewness		.129
Kurtosis		1.122
Std. Error of Kurtosis		.257
Minimum		0
Maximum		30
Percentiles (*(selected deciles and quartiles)*)	10	2.00
	25	5.00
	75	12.00
	90	16.00

Figure 5.2 Background knowledge test: frequency distribution



5.4 Reading comprehension indicators and variables

5.4.1 Introduction

As described in Chapter 4, a reading comprehension test with 17 short-answer questions referring to three texts was used in the main study. They were chosen after several rounds of preliminary study involving three financial experts and according to five criteria: text specificity, appropriateness, familiarity, provision of new knowledge, and explanation of concepts. Reducing the number of texts and reading comprehension test questions meant that the administration of the test battery would be less time consuming and more practical.

5.4.2 Aim

The test was intended to show how readers perform on reading comprehension of highly specific texts and link the results to their scores on tests of grammar knowledge and background knowledge, as well as their reports in the post-reading questionnaire on: a) self-assessed familiarity with text topics, b) interest in the texts, and c) perceived difficulty of texts.

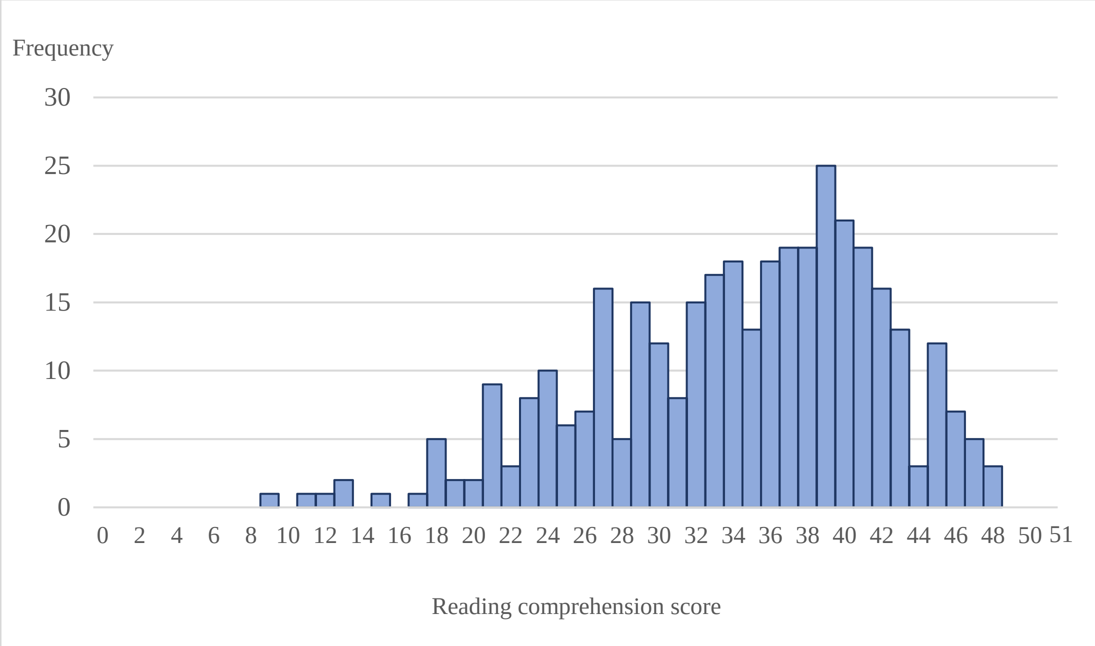
5.4.3 Descriptive statistics

The reading comprehension construct is based on 17-item reading comprehension test with a potential maximum score of 51 points. The short-answer questions were scored by using an ordinal scale with the following coding: 0 points (incorrect, irrelevant, don't know), 1 point (partly correct), 2 points (largely correct), and 3 points (correct). The composite index reading comprehension is distributed as shown in Table 5.3 and Figure 5.3. The curve is slightly left-skewed with a skewness coefficient of −.605 (standard error 0.129) and with a kurtosis coefficient of −.029 (standard error of 0.257). The values of this variable (i.e. test scores) are within the 9–48 point range out of 51, with a mean score of 34.25 and a standard deviation of 7.660. The median is 36 and mode 39. The tenth of students with the lowest scores achieved 23 or fewer points, whereas the 10% of students with the highest scores achieved 43 points or more.

Table 5.3 Reading comprehension test: descriptive statistics

Statistic		Reading Comprehension
n		358
Mean		34.25
Mean (%)		67.16%
Median		36.00
Mode		39.00
Std. Deviation		7.660
Skewness		-.605
Std. Error of Skewness		.129
Kurtosis		-.029
Std. Error of Kurtosis		.257
Minimum		9
Maximum		48
Percentiles ((*selected deciles and quartiles)*	10	23.00
	25	29.00
	75	40.00
	90	43.00

Figure 5.3 Reading comprehension test, construct: frequency distribution



5.4.4 Reading comprehension scores for individual texts

Table 5.4 presents descriptive statistics referring to reading comprehension scores for individual texts as well as the total results. Students performed best on reading comprehension of the SS text, achieving a 73.94% mean score. They achieved the lowest score on the BRR text at 57.53%, whilst the result for TS was in the middle at 68.44%. The mean reading comprehension score for all three texts together was 67.16%.

Looking at the distribution of scores across the range, it can be noted that the scores for BRR had a minimum score of 0% (0 out of 15 points), a maximum of 100% (15 out of 15 points), a mean reading comprehension score of 8.63 (that is 57.53%), SD at 3.45 and a coefficient of variation (CV) of 0.400, which indicates the highest variability of scores out of all the three texts. The standard deviation (SD) for the SS text was 3.36 (CV 0.252) and 2.97 for TS (CV 0.241). The maximum score for TS was 100% (all 18 points) and minimum 0% (no points), whereas for SS the minimum was 16.7% (3 points out of 18) and the maximum 100% (all 18 points).

In summary, the descriptive statistics of the reading comprehension scores for individual texts show that students found BRR the most difficult text, followed by TS, and SS the easiest. Section 5.9 presents univariate and bivariate analyses, comparing the reading comprehension of the texts with students' assessment in the post-reading questionnaire of how difficult they found each text, how familiar they were with its topic, and how interesting they found it.

Table 5.4 Descriptive statistics for reading comprehension scores of 3 texts (pts)

Text	n	Minimum	Maximum	Mean	Mean (%)	Median	Std. Deviation
SS (18 pts)	358	3	18	13.31	73.94%	14.00	3.36
BRR (15 pts)	358	0	15	8.63	57.53%	9.00	3.45
TS (18 pts)	358	0	18	12.32	68.44%	12.00	2.97
Total (51 pts)	358	9	48	34.25	67.16%	36.00	7.66

5.5 Post-reading questionnaire

5.5.1 Introduction and aim

The aim of the post-reading questionnaire was to obtain baseline data as well as students' self-reported data on topic familiarity, text difficulty and interest, which could be compared to students' scores on a background knowledge test and interpreted in relation to their reading comprehension scores. As pointed out in Chapter 2, a number of studies operationalized background knowledge primarily as either a) students' field of study (Alderson & Urquhart, 1983; Peretz & Shoham, 1990; Clapham, 1991; Ja'far 1992; Hill & Liu, 2012) or b) students' reports about their topic familiarity (Afflerbach 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Brantmeier, 2003; Lin, 2002; Nodoushan, 2003; Pulido, 2007; Leeser, 2007; Lee, 2007; McNeil, 2010; Eidswick, 2010; Rouhi & Ashgari, 2011; Lahuerta Martinez, 2013; Ashrafzadeh et al., 2015; Horiba & Fukaya, 2015). However, the field of study or familiarity self-reports in these studies were not combined with results obtained by testing background knowledge. This study aimed to overcome this shortcoming and introduce a more objective measure by contrasting familiarity self-reports with background knowledge test scores. In Section 5.4.5, I provide more information about the Self-reported Familiarity (SRF) construct, which is later used as a separate independent variable besides Grammar Knowledge and Background Knowledge in regression models addressing the research questions.

The questionnaire contained both closed and open questions (Appendix B-5). Answers to closed questions were marked on a seven-point scale. They were followed by open questions which aimed to obtain more detailed information. Students answered the questions in the questionnaire related to each text immediately after reading the text and answering the reading comprehension questions in order to limit the loss of information due to delays or even mix-ups that could have occurred if students had been asked to answer the questions about the three texts together at the end of the test administration.

5.5.2 Gender and reading habits

The post-reading questionnaire was filled in by 358 students, of whom 36.6% were male and 60.1% female (Table 5.5 and Appendix C-4); 3.4% of students did not answer this question.

Table 5.5 Post-reading questionnaire: gender

Variable	1 - Male	2 - Female	Missing
Gender	36.6%	60.1%	3.4%

Questions 2 and 3 in the questionnaire enquired about students' reading habits in English, both general and course-related, as shown in Table 5.6. The results for both questions are fairly similarly distributed, but the medians differed: 3.00 for general reading, and 4.00 for course-related reading. To get a clearer picture about reported reading habits, assessments for codes 1, 2, and 3 were collapsed on the one hand and 5, 6, and 7 on the other, leaving out the middle point (code 4), so that 1+2+3 stands for no or little reading, and 5+6+7 for considerable reading. The obtained figures show that 32.7% of students reported a considerable amount of general reading and 27.9% a considerable amount of course-related reading. On the other hand, 51.3% reported no or little amount of general reading and 44.4% no or little amount of course-related reading. The comparison of collapsed codes indicates that there were more students who assessed their reading habits as non-existent or low than considerable, both for general and course-related reading. However, the difference in student assessments of their general and course-related reading is small, suggesting that students claim to do a roughly similar amount of both general and course-related reading.

Table 5.6 Post-reading questionnaire: reading habits

Variable	1 Not at all	2	3	4	5	6	7 A great deal	Missing	Median
Do you read articles, magazines or books **on any topic** in English?	4.7%	24.0%	22.6%	14.5%	18.7%	8.1%	5.9%	1.4%	3.00
Do you read articles, magazines or books **for your coursework** at the university in English?	5.0%	18.2%	21.2%	26.3%	15.9%	10.3%	1.7%	1.4%	4.00

5.5.3 Difficulty, familiarity, and interest

After reading each of the three texts and answering reading comprehension questions, students answered the post-reading questionnaire. On a seven-point scale they assessed the level of text difficulty, their familiarity with the topic, and how interesting they found the text. They were also asked to answer two open questions: first, what they attribute the level of text difficulty to, and second, how their topic familiarity or lack of it affected their reading comprehension.

"Star Struck"

Tables 5.7 and 5.8 show students' ratings of a) difficulty, b) topic familiarity, c) the effect of familiarity and d) students' interest in the topic of the "Star Struck" text. The students found it relatively easy, as almost 50% of respondents rated text difficulty from 1–3. There were 20.1% sitting in the middle at point 4, and 29.0% found the text rather difficult, assessing it from 5–7. The students found the text fairly unfamiliar, with over 73.5% rating their familiarity from 1–3, resulting in a relatively low median of 2.00. When asked whether their topic familiarity affected their reading comprehension, about 50% reported a low effect (1–3), almost 30% a higher one (5–7), and 20.4% chose the medium effect (4). In terms of interest, 44.1% of students rated the SS text from 5–7 as quite interesting, 19.0% rated it of medium interest and just 26.5% rated it 1–3 as relatively uninteresting, with a median of 4.00. In comparison to the other two texts, students found "Star Struck" the easiest, the most interesting, and having the lowest effect of topic familiarity.

Table 5.7 Post-reading questionnaire: "Star Struck", difficulty

Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	Missing	Median
How **difficult** did you find the text? SS	6.1 %	21.5 %	23.2 %	20.1 %	18.7 %	8.9 %	1.4 %	0.0 %	3.00

Table 5.8 Post-reading questionnaire: "Star Struck", familiarity and interest

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	Missing	Median
How **familiar** were you with the topic before you read the text? SS	32.4 %	24.3 %	16.8 %	12.0 %	11.2 %	2.5 %	.3 %	0.6 %	2.00
How much did this affect your reading comprehension? SS	10.6 %	18.4 %	20.9 %	20.4 %	16.8 %	9.2 %	3.1 %	0.6 %	3.00
How **interesting** did you find the text? SS	2.8 %	9.5 %	14.2 %	19.0 %	19.8 %	20.9 %	3.4 %	10.3 %	4.00

"Bonds that Rock and Roll"

Students found the text "Bonds that Rock and Roll" more difficult than "Star Struck," with 57.5% rating the difficulty 5–7, 17.0% opting for the medium point 4 and 24.6% rating it as relatively easy (1–3; Table 5.9). In terms of familiarity, 81.8% rated the topic as fairly unfamiliar (1–3) and only 9.2% rated it as relatively familiar (5–7). Similar to the SS text, 49.8% perceived no or low effect of familiarity on their reading comprehension (1–3), 31.3% reported a higher effect of familiarity (5–7), and 18.2% perceived a medium effect. Finally, the text was rated as relatively interesting (5–7) by 30.7%, 16.2% assessed it at a medium level of interest (4), and 41.9% assessed it at a lower degree of interest (1–3).

In comparison to the other texts, BRR was found to be a relatively difficult text and the most unfamiliar of the three, although the perceived familiarity was reported to have a medium effect (Table 5.10). In fact, the difficulty level was rated as just below TS, with a median of 5.00, and over 57% of respondents rating the difficulty from 5–7. The perceived topic familiarity with BRR was rated lowest of the three texts by 81.8% of respondents, with a median of 2.00. However, the effect of familiarity was at almost the same level as for the SS text, as 49.8% claimed a low effect of familiarity (1–3) and 31.3% a high effect (5–7). As for the level of interest, the BRR text was rated overall as less interesting than SS and more than TS.

Table 5.9 Post-reading questionnaire: "Bonds that Rock and Roll", difficulty

Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	Missing	Median
How **difficult** did you find the text? BRR	.6 %	7.8 %	16.2 %	17.0 %	26.0 %	25.4 %	6.1 %	.8 %	5.00

Table 5.10 Post-reading questionnaire: "Bonds that Rock and Roll", familiarity and interest

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	Missing	Median
How **familiar** were you with the topic before you read the text? BRR	30.7 %	33.2 %	17.9 %	8.4 %	6.7 %	2.2 %	.3 %	.6 %	2.00
How much did this affect your reading comprehension? BRR	8.7 %	20.7 %	20.4 %	18.2 %	16.5 %	9.8 %	5.0 %	.8 %	3.00
How **interesting** did you find the text? BRR	7.5 %	15.1 %	19.3 %	16.2 %	17.3 %	11.7 %	1.7 %	11.2 %	4.00

“Taking Stock”

Tables 5.11 and 5.12 show the student assessment of the “Taking Stock” text. 57.5% of students found it rather difficult (5–7), 23.5% assessed it as medium difficulty (4), and 18.8% thought it was relatively easy (1–3). As for familiarity, 57.5% of students reported the topic as unfamiliar (1–3), 17.9% perceived a medium degree of familiarity (4), and 24.3% thought the topic was quite familiar (5–7). 41.6% of students felt that they were relatively unaffected by familiarity (1–3), whereas 37,6% claimed they were very affected by it (5–7), and 19.3% reported a medium effect. Responding to the question about interest, 22.3% of students found the text relatively interesting (5–7), 19.8% felt a medium degree of interest, and 46.7% rated it 1–3 or not particularly interesting.

A brief comparison of “Taking Stock” to the other two texts suggests that it was found to be the most difficult of the three texts albeit not by a statistically significant margin: TS and BRR were both with medians of 5.00 (Tables 5.10 and 5.12). Its topic was perceived the most familiar of all, and the familiarity was reported to have the strongest effect on reading comprehension, shown by the mean of 3.92, compared to 3.54 for SS and 3.63 for BRR. However, TS was found to be the least interesting of the three texts, shown by the mean of 3.47.

Table 5.11 Post-reading questionnaire: “Taking Stock”, difficulty

Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	Missing	Median
How **difficult** did you find the text? TS	1.4 %	5.9 %	11.5 %	23.5 %	27.9 %	22.3 %	7.3 %	.3 %	5.00

Table 5.12 Post-reading questionnaire: "Taking Stock", familiarity and interest

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	Missing	Median
How **familiar** were you with the topic before you read the text? TS	13.1 %	20.9 %	23.5 %	17.9 %	15.6 %	8.1 %	.6 %	.3 %	3.00
How much did this affect your reading comprehension? TS	5.6 %	15.1 %	20.9 %	19.3 %	20.1 %	12.8 %	4.7 %	1.4 %	4.00
How **interesting** did you find the text? TS	9.2 %	14.0 %	23.5 %	19.8 %	14.2 %	6.1 %	2.0 %	11.2 %	3.00

5.5.4 Comparison of texts according to perceived difficulty, familiarity, effect of familiarity, and interest

Tables 5.13, 5.14, 5.15, and 5.16 illustrate comparisons of the three texts according to student ratings of text difficulty, topic familiarity, the effect of familiarity, and interest in the post-reading questionnaire. In order to make the picture clearer, the ratings 1, 2, and 3 were collapsed into 1–3 (Low level), and 5, 6, and 7 into 5–7 (High level) because they represent clear ends of the scale, whereas the middle point 4 was kept unchanged.

a) Difficulty

Students perceived SS to be the easiest text, with 50.8% of students rating it 1–3, whereas both TS and BRR were found to be more difficult and at a similar difficulty level (57.5% rating it 5–7).

Table 5.13 Post-reading questionnaire: perceived difficulty, all texts

Perceived Difficulty	1–3 Low	4	5–7 High	Missing	Median
SS	50.8%	20.1%	29.1%	0.0%	3.00
BRR	24.6%	17.0%	57.5%	0.8%	5.00
TS	18.7%	23.5%	57.5%	0.3%	5.00

b) Familiarity

The familiarity assessments results suggested that in general the students thought they were not particularly familiar with the text topics (medians between 2.00 and 3.00). The least familiar of all texts was BRR, with 81.8% rating it 1–3 (low familiarity). In terms of low familiarity SS came second with 73.5% rating it 1–3. TS came out as the most familiar of all texts, although still 57.5% assessed it 1–3 (low familiarity).

Table 5.14 Post-reading questionnaire: perceived familiarity, all texts

Perceived Familiarity	1–3 Low	4	5–7 High	Missing	Median
SS	73.5%	12.0%	14.0%	0.6%	2.00
BRR	81.8%	8.4%	9.2%	0.6%	2.00
TS	57.5%	17.9%	24.3%	0.3%	3.00

c) Effect of familiarity

When rating how much effect familiarity had on their reading comprehension, students gave ratings that were relatively similar among the texts. They reported that familiarity with the topic of TS had the highest effect, with around 40% claiming a high effect and almost 40% a low effect and the rest sitting in between or missing (with median of 4.0). As for BRR and SS, most students reported an effect below the midpoint (with medians of 3.00).

Table 5.15 Post-reading questionnaire: perceived effect of familiarity, all texts

Perceived Effect of Familiarity	1–3 Low	4	5–7 High	Missing	Median
SS	50.0%	20.4%	29.1%	0.6%	3.00
BRR	49.7%	18.2%	31.3%	0.8%	3.00
TS	41.6%	19.3%	37.7	1.4%	4.00

d) Interest

Of the three texts, SS was found to be the most interesting with a mean of 4.34, followed by BRR at 3.70 and TS at 3.47. This variable had the highest number of missing answers (10.3%–11.2%).

Table 5.16 Post-reading questionnaire: interest, all texts

Interest	1–3 Low	4	5–7 High	Missing	Median
SS	26.5%	19.0%	44.1%	10.3%	4.00
BRR	41.9%	16.2%	30.7%	11.2%	4.00
TS	46.6%	19.8%	22.3%	11.2%	3.00

A comparison of student ratings for all three texts and all parameters leads to the following observations:

1. Of the four studied concepts (Tables 5.13-5.16), and relatively speaking, the highest median was reported for difficulty and the lowest average level was reported for familiarity.
2. The level of familiarity was reported to have a medium effect on reading comprehension; similarly, interest was rated at the middle level for two texts and above it for one.
3. The text that was perceived as the easiest (SS), though relatively unfamiliar, was found to be most interesting and a medium effect of familiarity was reported.
4. The text that was perceived to be the most difficult (TS) was also perceived as the most familiar of three texts but least interesting as well. Students reported the effect of familiarity in this case to be the strongest.

In order to interpret the results of the post-reading questionnaire, student ratings needed to be paired up first to the responses to open questions in the post-reading questionnaire (in section 5.4.6), and second, to the results of the testing study. The aim was to make the following comparisons:

a) Students' self-reported familiarity ratings vs. their scores on the background knowledge test and reading comprehension test;

b) Students' text difficulty ratings vs. text readability measured by the Flesch Kincaid readability index and Fog index presented in Chapter 4;
c) Students' text difficulty ratings vs. their scores on the reading comprehension test;
d) Students' interest ratings vs. their scores on the reading comprehension test.

The comparison and correlation of these parameters will be addressed in section 5.8. Descriptive statistics for Self-reported Familiarity (SRF), a construct created from familiarity scores for SS, BRR and TS texts, are presented in section 5.4.5.

5.5.5 Descriptive statistics for the Self-reported Familiarity construct

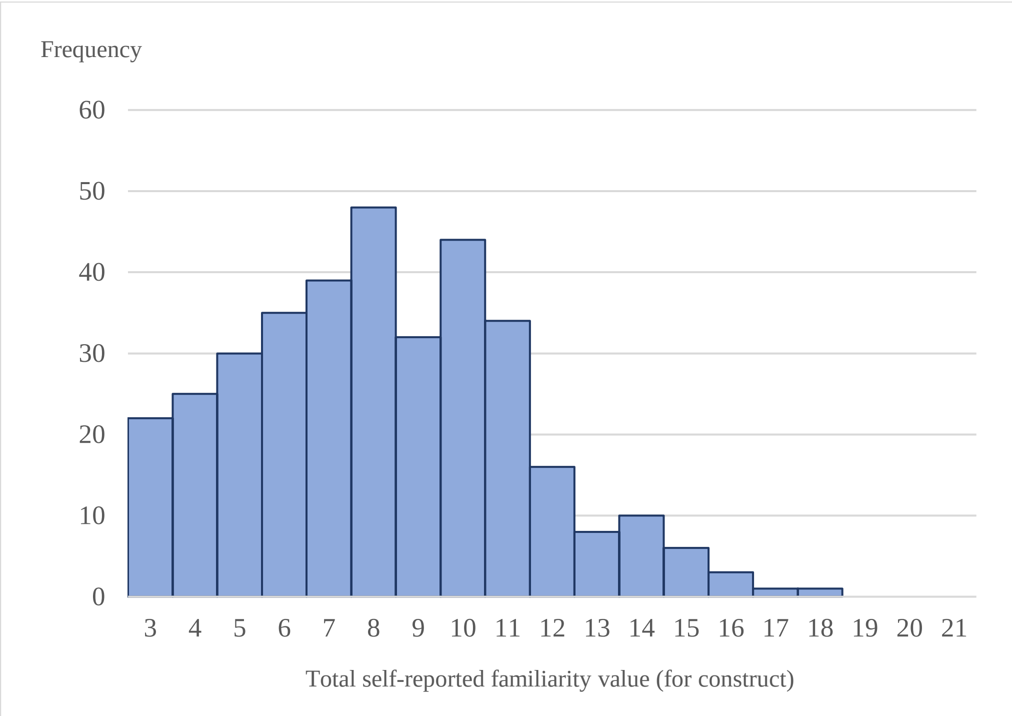
The Self-reported Familiarity (SRF) construct is based on self-assessment of research participants – they answered the question "How familiar were you with the topic before you read the text?" using a 7-point scale for the following three topics: SS, BRR, and TS (see Tables 5.8, 5.10, and 5.12 for individual distributions of familiarity). Thus, a potential minimum score was 3 (3 times 1) and a potential maximum score was 21 (3 times 7). For reliability analysis see Section 4.7 in Chapter 4.

The composite index Self-reported Familiarity is distributed as shown in Table 5.17 and Figure 5.4. The curve is slightly right-skewed with a skewness coefficient of .314 (standard error 0.130) and with a kurtosis coefficient of -.241 (standard error of 0.259). The values of this variable (i.e. combined scale values) are within the 3–18 value range out of 21, with a mean score of 8.18 and a standard deviation of 3.097. The median and mode are both 8.00. The tenth of students with the lowest value self-reported familiarity were those with a value of 4 or less, whereas the top 10% of students were with a combined value of 12 or more.

Table 5.17 Self-reported Familiarity: descriptive statistics

Statistic		Self-reported familiarity construct
n		354
Mean		8.18
Std. Error of Mean		.165
Median		8.00
Mode		8.00
Std. Deviation		3.097
Skewness		.314
Std. Error of Skewness		.130
Kurtosis		-.241
Std. Error of Kurtosis		.259
Minimum		3
Maximum		18
Percentiles (*(selected deciles and quartiles)*)	10	4.00
	25	6.00
	75	10.00
	90	12.00

Figure 5.4 Self-reported Familiarity: frequency distribution



5.5.6 Answers to open questions

The questionnaire contained two open questions that followed student ratings of text difficulty and the effect of text familiarity:

- What do you attribute the level of text difficulty to?
- In what way did your familiarity with the topic or lack of it show in your reading comprehension?

It was expected that the answers to these two questions would shed more light into understanding the difficulties students had in reading comprehension, and in what way topic familiarity played a part in this.

5.5.7 Coding of answers to open questions

The number of answers given by students is summed up in Table 5.18. Some students provided more than one answer, whilst others gave no answer at all. For example, the question about the sources of difficulty for each text was answered by 44.4%–56.7% of students providing one source or reason, 19.6%–24% respondents providing two reasons, 19.6%–24% giving three reasons, and two students listing four reasons (for BRR text only). However, 21.2%–27.4% of the students did not answer the question about the sources of text difficulty. The open question about how familiarity showed in readers' comprehension was answered by 52.5%–59.8% giving one effect, 4.5%–11.5% giving two effects, and two students (0.6%) reporting three effects. 34.4%–41.9% of students didn't answer the question.

Table 5.18 Open answers: difficulty and effect of familiarity

Difficulty: Open Answers	Star Struck		Bonds That Rock and Roll		Taking Stock	
Single coding	203	56.7%	159	44.4%	196	54.7%
Double coding	72	20.1%	86	24.0%	70	19.6%
Triple coding	7	2.0%	13	3.6%	10	2.8%
Quadruple coding	0	0.0%	2	0.6%	0	0.0%
No answer provided	76	21.2%	98	27.4%	82	22.9%
Number of respondents	358	100.0%	358	100.0%	358	100.0%
Effect of Familiarity: Open Answers	**Star Struck**		**Bonds That Rock and Roll**		**Taking Stock**	
Single coding	214	59.8%	192	53.6%	188	52.5%
Double coding	19	5.3%	16	4.5%	41	11.5%
Triple coding	2	0.6%	0	0.0%	0	0.0%
No answer provided	123	34.4%	150	41.9%	129	36.0%
Number of respondents	358	100.0%	358	100.0%	358	100.0%

The answers to both open questions were examined, and the most frequent answers were used in developing a list of codes that was used for subsequent answer coding. This was done in three rounds to keep the answers under *Other* as low as possible. The final coding list for difficulty comprises 16 codes, and the coding list for effect of familiarity consists of 13 codes as shown in Table 5.19.

Table 5.19 Coding of answers to open questions: sources of difficulty and effects of familiarity

Sources of Difficulty	Effects of Familiarity
1. Vocabulary 2. Unknown vocabulary 3. Known vocabulary 4. Specialist vocabulary 5. Non-specialist vocabulary 6. Difficult vocabulary	1. Faster reading 2. Slower reading 3. Easier to understand 4. More difficult to understand 5. Easier to make connections: inferring, predicting
7. Difficult text 8. Easy text 9. Long text 10. Short text	6. Known/easy vocabulary 7. Unknown/difficult vocabulary 8. Interesting topic 9. Uninteresting topic 10. Familiar topic 11. Unfamiliar topic
11. Topic 12. Familiar topic 13. Unfamiliar topic 14. Interesting topic 15. Uninteresting topic	12. It didn't show
16. Other	13. Other

A closer look at the 16 codes related to the answers to questions about difficulty shows that, apart from the code *Other*, they could be divided into three main groups referring to: a) vocabulary, b) text, and c) topic. The 13 codes for answers to questions about the effects of familiarity could be broken down into five key areas of reference: a) reading speed, b) understanding, c) vocabulary, d) topic, and e) no effect.

The coding of answers to the open questions revealed that students' answers provided different degrees of detail. For instance, students cited simply *Vocabulary* or *Topic* as a source of difficulty, without indicating exactly what the problem was (Table 5.19). Conversely, other answers defined vocabulary as *difficult*, *unknown*, or *specialist*, as well as their opposites *known* and *non-specialist*. Similarly, some students reported *Topic* as a source of difficulty, whereas others further defined it as *familiar* or *unfamiliar* and *interesting* or *uninteresting*.

When asked about how topic familiarity or lack of it showed in their reading comprehension, the highest number of answers referred to a) reading speed or b) ease or difficulty of comprehension and making connections, including inferring or predicting. However, there were a number of respondents who linked effects of familiarity to c) the topic; they reported that their level of familiarity showed in finding the topic as more or less interesting (Table 5.19, codes *Interesting topic*, *Uninteresting topic*). Some students just reiterated their familiarity or lack of it by saying: *Familiar topic*, or *Unfamiliar topic*, although this was not the answer to the question but just a repetition of it. This may have happened because students either misunderstood the question, or may not have been aware how their familiarity showed in reading comprehension. In the first round of coding these items were categorized as *Irrelevant* because they failed to answer to the question, but eventually they were kept as separate codes because the irrelevance of answers may also be indicative of students' awareness.

The coding of answers also showed that some answers may overlap. For instance, in relation to the "Star Struck" text 46.9% of students reported vocabulary as a reason for the level of text difficulty (Table 5.20). Among these answers, there were different degrees of detail. So, 2.5% answers were just *Vocabulary,* without indicating what exactly made it difficult. 14% answers referred to *Unknown vocabulary*, 3.1% *Specialist vocabulary*, 8.9% *Difficult vocabulary*, 7.8% *Known vocabulary*, 10.6% *Non-specialist vocabulary.* Even though there may be overlap among these answers, the answers were listed under separate codes. For instance, a respondent may have reported a string like the following: 'difficult unknown specialist vocabulary', which points to the interpretation that specialist vocabulary is unknown to them and is therefore difficult.

5.5.8 Perceived sources of text difficulty

Table 5.20 shows students' answers to the question regarding the reasons for text difficulty referring to the three texts. It gives both the frequency of answers and the percentage of respondents who gave that answer. The percentage exceeds 100% because there were students who gave more than one answer. For all three texts the most common answer was *Unknown vocabulary,* with 24.9% for TS, 18.4% for BRR, and 14.0% for SS, thus clearly showing the most important source of difficulty as perceived by students. To further distil the information, the answers that had a degree of overlap were collapsed as shown in Table 5.21. For instance, *Vocabulary, Unknown vocabulary, Specialist vocabulary, and Difficult vocabulary* were collapsed. This presentation clearly highlights vocabulary as a prevailing source of perceived difficulty, with 50.0% answers for TS, 35.7% for BRR, and 28.5% for SS. The second most frequent answer was *Long text* for BRR with 21.8%, followed by *Difficult text* with 13.1% for TS and 10.6% for BRR.

Table 5.20 Perceived sources of text difficulty by frequency

What do you attribute the level of **difficulty** to?	**Star Struck**		What do you attribute the level of **difficulty** to?	**Bonds That Rock and Roll**		What do you attribute the level of **difficulty** to?	**Taking Stock**	
	Frequency	Percent of all respondents		Frequency	Percent of all respondents		Frequency	Percent of all respondents
8 Easy text	65	18.2%	9 Long text	78	21.8%	2 Unknown vocabulary	89	24.9%
2 Unknown vocabulary	50	14.0%	2 Unknown vocabulary	66	18.4%	7 Difficult text	47	13.1%
5 Non-specialist vocabulary	38	10.6%	7 Difficult text	38	10.6%	6 Difficult vocabulary	37	10.3%
6 Difficult vocabulary	32	8.9%	13 Unfamiliar topic	31	8.7%	16 Other	37	10.3%
7 Difficult text	32	8.9%	8 Easy text	30	8.4%	4 Specialist vocabulary	35	9.8%
3 Known vocabulary	28	7.8%	6 Difficult vocabulary	28	7.8%	13 Unfamiliar topic	21	5.9%
10 Short text	18	5.0%	4 Specialist vocabulary	22	6.1%	8 Easy text	19	5.3%
13 Unfamiliar topic	18	5.0%	16 Other	22	6.1%	1 Vocabulary	18	5.0%
14 Interesting topic	16	4.5%	5 Non-specialist vocabulary	13	3.6%	12 Familiar topic	14	3.9%
16 Other	16	4.5%	1 Vocabulary	12	3.4%	5 Non-specialist vocabulary	10	2.8%
12 Familiar topic	14	3.9%	3 Known vocabulary	10	2.8%	11 Topic	9	2.5%
11 Topic	12	3.4%	11 Topic	9	2.5%	15 Uninteresting topic	9	2.5%
4 Specialist vocabulary	11	3.1%	14 Interesting topic	8	2.2%	9 Long text	8	2.2%
1 Vocabulary	9	2.5%	12 Familiar topic	5	1.4%	10 Short text	5	1.4%
15 Uninteresting topic	7	2.0%	15 Uninteresting topic	5	1.4%	3 Known vocabulary	4	1.1%
9 Long text	2	0.6%	10 Short text	1	0.3%	14 Interesting topic	4	1.1%
No answer provided	76	21.2%	No answer provided	98	27.4%	No answer provided	82	22.9%
Total	**444**		**Total**	**476**		**Total**	**448**	

Table 5.21 Most common sources of text difficulty

Star Struck	**28.5%** - Vocabulary, Unknown vocabulary, Specialist vocabulary, Difficult vocabulary 18.4% - Known vocabulary, Non-specialist vocabulary 18.2% - Easy text
Bonds that Rock and Roll	**35.7%** - Vocabulary, Unknown vocabulary, Specialist vocabulary, Difficult vocabulary 21.8% - Long text 10.6% - Difficult text
Taking Stock	**50.0%** - Vocabulary, Unknown vocabulary, Specialist vocabulary, Difficult vocabulary 13.1% - Difficult text 10.3% - Other

The next step in analysing answers to open questions was to carry out bivariate analyses for each text by cross-tabulating answer frequencies with the number of students who gave each text a particular difficulty rate. The intention was to see whether there were any common trends in answers to open questions provided by students who found the text either very difficult or very easy. Finally, the aim was to make a comparison of results between the texts.

Bivariate analyses of the sources of difficulty for the three texts are presented in Tables 5.22, 5.23, and 5.24. In analyses of SS and BRR *No answers provided* had the highest frequency (i.e. 98 for BRR and 76 for SS), whereas it was second most frequent for TS (i.e. 82). Non-respondents were evenly dispersed as for their difficulty rating. In contrast, there was a stronger focus of answers for TS, with most difficulty ratings between 4 and 6. This suggests that those students who found TS more difficult also failed to give their answers about why they did so.

Cross-tabulation of difficulty ratings and attributions: "Star Struck"

The following observations can be made in relation to cross-tabulations of difficulty ratings and difficulty sources for the Star Struck text as presented in Table 5.22.

- The first most frequent answer was *Easy text*, given by 65 students, and 54 of these (or 83.1%) rated SS as a low level of difficulty (1–3), with a median of 2.00. In giving the answer to the open question the students just reiterated their assessment, just like they did for *Difficult text* (71.9% rated SS as a medium to high level of difficulty, median 4.00).

- The second most frequent answer *Unknown vocabulary* was provided by 50 students, and 46 of these (or 92.0%) rated SS as medium level of difficulty (3–5), with a median of 4.00.
- The third most frequent answer *Non-specialist vocabulary* was given by 38 students, and 28 of these (or 73.7%) rated the text as relatively easy (2–3), with a median of 2.00).

Table 5.22 Bivariate analysis of the "Star Struck" text

5	**How difficult did you find the text? Star Struck**							**Total**	**Median**
	1	**2**	**3**	**4**	**5**	**6**	**7**		
8 Easy text	11	31	12	7	3	1	0	65	2.00
	16.9%	47.7%	18.5%	10.8%	4.6%	1.5%	0.0%	100.0%	
2 Unknown vocabulary	0	0	11	20	15	4	0	50	4.00
	0.0%	0.0%	22.0%	40.0%	30.0%	8.0%	0.0%	100.0%	
5 Non-specialist vocabulary	6	17	11	1	3	0	0	38	2.00
	15.8%	44.7%	28.9%	2.6%	7.9%	0.0%	0.0%	100.0%	
6 Difficult vocabulary	1	1	7	11	9	3	0	32	4.00
	3.1%	3.1%	21.9%	34.4%	28.1%	9.4%	0.0%	100.0%	
7 Difficult text	1	2	5	8	7	7	2	32	4.50
	3.1%	6.3%	15.6%	25.0%	21.9%	21.9%	6.3%	100.0%	
3 Known vocabulary	2	12	12	1	1	0	0	28	2.00
	7.1%	42.9%	42.9%	3.6%	3.6%	0.0%	0.0%	100.0%	
10 Short text	4	11	2	0	0	1	0	18	2.00
	22.2%	61.1%	11.1%	0.0%	0.0%	5.6%	0.0%	100.0%	
13 Unfamiliar topic	0	0	4	5	6	2	1	18	4.50
	0.0%	0.0%	22.2%	27.8%	33.3%	11.1%	5.6%	100.0%	
14 Interesting topic	0	9	6	1	0	0	0	16	2.00
	0.0%	56.3%	37.5%	6.3%	0.0%	0.0%	0.0%	100.0%	
16 Other	0	4	3	3	2	4	0	16	4.00
	0.0%	25.0%	18.8%	18.8%	12.5%	25.0%	0.0%	100.0%	
12 Familiar topic	0	3	7	3	1	0	0	14	3.00
	0.0%	21.4%	50.0%	21.4%	7.1%	0.0%	0.0%	100.0%	
11 Topic	0	5	4	1	1	1	0	12	3.00
	0.0%	41.7%	33.3%	8.3%	8.3%	8.3%	0.0%	100.0%	
4 Specialist vocabulary	0	0	3	4	4	0	0	11	4.00
	0.0%	0.0%	27.3%	36.4%	36.4%	0.0%	0.0%	100.0%	
1 Vocabulary	0	1	2	3	3	0	0	9	4.00
	0.0%	11.1%	22.2%	33.3%	33.3%	0.0%	0.0%	100.0%	
15 Uninteresting topic	0	0	1	2	2	1	1	7	5.00
	0.0%	0.0%	14.3%	28.6%	28.6%	14.3%	14.3%	100.0%	
9 Long text	0	0	1	1	0	0	0	2	3.50
	0.0%	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%	100.0%	
Total	**25**	**96**	**91**	**71**	**57**	**24**	**4**	**368**	**3.00**
	6.8%	**26.1%**	**24.7%**	**19.3%**	**15.5%**	**6.5%**	**1.1%**		

Cross-tabulation of text difficulty ratings and attributions: "Bonds that Rock and Roll"

The following observations can be made with respect to cross-tabulations of difficulty ratings and difficulty sources for the "Bonds that Rock and Roll" text shown in table 5.23:

- The first most frequent answer *Long text* was given by 77 students, and 68 of these (or 88.3%) assessed it as medium and difficult (4–7), with a median of 5.00.
- The second most common answer *Unknown vocabulary* was given by 66 students, and 59 of these (or 89.4%) rated text difficulty medium and high (4–7), with a median of 5.00.
- The third most frequent answer *Difficult text* was given by 38 students, and 33 of these (or 86.8%) rated the text as relatively difficult (5–7), with a median of 6.00. Basically, students repeated their assessment in their answers to the open question.

Table 5.23 Bivariate analysis of the "Bonds that Rock and Roll" text

What do you attribute the level of difficulty to?	How difficult did you find the text? Bonds That Rock and Roll							Total	Median
	1	2	3	4	5	6	7		
9 Long text	0	0	9	18	21	17	12	77	5.00
	0.0%	0.0%	11.7%	23.4%	27.3%	22.1%	15.6%	100.0%	
2 Unknown vocabulary	0	0	7	14	23	18	4	66	5.00
	0.0%	0.0%	10.6%	21.2%	34.8%	27.3%	6.1%	100.0%	
7 Difficult text	0	0	1	4	10	14	9	38	6.00
	0.0%	0.0%	2.6%	10.5%	26.3%	36.8%	23.7%	100.0%	
13 Unfamiliar topic	0	0	2	6	14	8	0	31	5.00
	0.0%	0.0%	6.5%	19.4%	45.2%	25.8%	0.0%	100.0%	
8 Easy text	2	10	13	3	1	1	0	30	3.00
	6.7%	33.3%	43.3%	10.0%	3.3%	3.3%	0.0%	100.0%	
6 Difficult vocabulary	0	0	0	1	8	13	6	28	6.00
	0.0%	0.0%	0.0%	3.6%	28.6%	46.4%	21.4%	100.0%	
4 Specialist vocabulary	0	0	2	1	10	9	0	22	5.00
	0.0%	0.0%	9.1%	4.5%	45.5%	40.9%	0.0%	100.0%	
16 Other	1	2	4	4	5	3	3	22	4.50
	4.5%	9.1%	18.2%	18.2%	22.7%	13.6%	13.6%	100.0%	
5 Non-specialist vocabulary	1	4	7	1	0	0	0	13	3.00
	7.7%	30.8%	53.8%	7.7%	0.0%	0.0%	0.0%	100.0%	
1 Vocabulary	0	0	2	2	3	3	2	12	5.00
	0.0%	0.0%	16.7%	16.7%	25.0%	25.0%	16.7%	100.0%	
3 Known vocabulary	0	4	4	2	0	0	0	10	3.00
	0.0%	40.0%	40.0%	20.0%	0.0%	0.0%	0.0%	100.0%	
11 Topic	0	1	0	1	1	5	1	9	6.00
	0.0%	11.1%	0.0%	11.1%	11.1%	55.6%	11.1%	100.0%	
14 Interesting topic	0	2	4	2	0	0	0	8	3.00
	0.0%	25.0%	50.0%	25.0%	0.0%	0.0%	0.0%	100.0%	
12 Familiar topic	0	4	1	0	0	0	0	5	2.00
	0.0%	80.0%	20.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
15 Uninteresting topic	0	0	0	1	1	3	0	5	6.00
	0.0%	0.0%	0.0%	20.0%	20.0%	60.0%	0.0%	100.0%	
10 Short text	1	0	0	0	0	0	0	1	1.00
	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
Total	**5**	**27**	**56**	**60**	**97**	**95**	**36**	**377**	**5.00**
	1.3%	**7.2%**	**14.9%**	**15.9%**	**25.7%**	**25.2%**	**9.5%**		

Cross-tabulation of text difficulty ratings and attributions: "Taking Stock"

The following observations can be made with regard to cross-tabulations of difficulty ratings and difficulty sources for the "Taking Stock" text as presented in Table 5.24.

- The most frequent answer *Unknown vocabulary* was given by 89 students, and 75 of these (or 84.3%) rated text difficulty as medium to high (4–6), with a median of 5.00.
- The second most frequent answer *Difficult text* was given by 47 students, 37 of whom (or 78.7%) rated TS as relatively difficult (5–7), with a median of 5.00. Again students here reiterate their ratings in their answers to open questions.
- The third most frequent answers were *Difficult vocabulary* and *Other,* and they were given by 37 students each. Of 37 participants who gave *Difficult vocabulary* answer, 28 (or 75.7%) rated the text as relatively difficult (5–7), with a median of 5.00. Of 37 participants who gave *Other* answer, 24 (or 64.9%) rated the text as fairly difficult (5-6), with a median of 5.00.
- The fifth most frequent answer *Specialist vocabulary* was given by 35 students, and 25 of these (or 71.4%) assessed the text as relatively difficult (5–7), with a median of 5.00.

Table 5.24 Bivariate analysis of the "Taking Stock" text

What do you attribute the level of **difficulty** to?	**How difficult** did you find the text? Taking Stock							**Total**	**Median**
	1	**2**	**3**	**4**	**5**	**6**	**7**		
2 Unknown vocabulary	0	2	7	22	34	19	5	89	5.00
	0.0%	2.2%	7.9%	24.7%	38.2%	21.3%	5.6%	100.0%	
7 Difficult text	0	1	2	7	16	13	8	47	5.00
	0.0%	2.1%	4.3%	14.9%	34.0%	27.7%	17.0%	100.0%	
6 Difficult vocabulary	0	1	4	4	10	13	5	37	5.00
	0.0%	2.7%	10.8%	10.8%	27.0%	35.1%	13.5%	100.0%	
16 Other	2	0	6	5	15	9	0	37	5.00
	5.4%	0.0%	16.2%	13.5%	40.5%	24.3%	0.0%	100.0%	
4 Specialist vocabulary	0	2	3	5	9	13	3	35	5.00
	0.0%	5.7%	8.6%	14.3%	25.7%	37.1%	8.6%	100.0%	
13 Unfamiliar topic	0	0	0	6	6	6	3	21	5.00
	0.0%	0.0%	0.0%	28.6%	28.6%	28.6%	14.3%	100.0%	
8 Easy text	1	6	4	3	4	1	0	19	3.00
	5.3%	31.6%	21.1%	15.8%	21.1%	5.3%	0.0%	100.0%	
1 Vocabulary	0	0	2	4	8	4	0	18	5.00
	0.0%	0.0%	11.1%	22.2%	44.4%	22.2%	0.0%	100.0%	
12 Familiar topic	0	4	4	6	0	0	0	14	3.00
	0.0%	28.6%	28.6%	42.9%	0.0%	0.0%	0.0%	100.0%	
5 Non-specialist vocabulary	0	3	5	1	0	1	0	10	3.00
	0.0%	30.0%	50.0%	10.0%	0.0%	10.0%	0.0%	100.0%	
11 Topic	1	0	1	2	4	1	0	9	5.00
	11.1%	0.0%	11.1%	22.2%	44.4%	11.1%	0.0%	100.0%	
15 Uninteresting topic	0	1	0	0	4	4	0	9	5.00
	0.0%	11.1%	0.0%	0.0%	44.4%	44.4%	0.0%	100.0%	
9 Long text	0	0	0	4	1	1	2	8	4.00
	0.0%	0.0%	0.0%	50.0%	12.5%	12.5%	25.0%	100.0%	
10 Short text	0	1	2	1	0	1	0	5	3.00
	0.0%	20.0%	40.0%	20.0%	0.0%	20.0%	0.0%	100.0%	
3 Known vocabulary	0	2	2	0	0	0	0	4	2.00
	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
14 Interesting topic	0	1	1	2	0	0	0	4	3.00
	0.0%	25.0%	25.0%	50.0%	0.0%	0.0%	0.0%	100.0%	
Total	**4**	**24**	**43**	**72**	**111**	**86**	**26**	**366**	**5.00**
	1.1%	**6.6%**	**11.7%**	**19.7%**	**30.3%**	**23.5%**	**7.1%**		

5.5.9 Perceived effects of familiarity

Table 5.25 presents student answers to the question about the effect of their familiarity or lack of it on their reading comprehension. The table provides the frequency of answers and the percentage of respondents who gave that answer. The percentage exceeds 100% because there were students who gave more than one answer. The first most frequent answer for TS and BRR was *More difficult to understand*, with 16.8% for TS and 14.0% for BRR. *It didn't show* was the most frequent answer for SS with 15.1% of answers. This was followed by *Easier to understand* in the second place with 10.6% for SS and 13.7% for TS and (only 6.1% for BRR). The second most frequent answer for BRR was *Unknown/difficult vocabulary*, which was also the third most common for TS with 9.8% of answers. *No answer provided* accounted for the highest percentage in relation to all three texts (41.9% for BRR, 36.0% for TS, and 34.4% for SS).

Table 5.25 Perceived effects of familiarity

In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	Star Struck		In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	Bonds That Rock and Roll		In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	Taking Stock	
	Frequency	Percent of all respondents		Frequency	Percent of all respondents		Frequency	Percent of all respondents
10 It didn't show	54	15.1%	4 More difficult to understand	50	14.0%	4 More difficult to understand	60	16.8%
3 Easier to understand	38	10.6%	6 Unknown/difficult vocabulary	37	10.3%	3 Easier to understand	49	13.7%
4 More difficult to understand	35	9.8%	10 It didn't show	31	8.7%	6 Unknown/difficult vocabulary	35	9.8%
6 Unknown/difficult vocabulary	26	7.3%	13 Other	27	7.5%	10 It didn't show	23	6.4%
2 Slower reading	18	5.0%	3 Easier to understand	22	6.1%	7 Known/easy vocabulary	22	6.1%
11 Unknown topic	17	4.7%	2 Slower reading	16	4.5%	13 Other	19	5.3%
13 Other	17	4.7%	9 Uninteresting topic	10	2.8%	5 Easier to make connections	18	5.0%
8 Interesting topic	15	4.2%	5 Easier to make connections	8	2.2%	12 Known topic	17	4.7%
7 Known/easy vocabulary	13	3.6%	8 Interesting topic	7	2.0%	2 Slower reading	11	3.1%
12 Known topic	9	2.5%	11 Unknown topic	6	1.7%	9 Uninteresting topic	8	2.2%
5 Easier to make connections	7	2.0%	12 Known topic	6	1.7%	1 Faster reading	4	1.1%
9 Uninteresting topic	5	1.4%	7 Known/easy vocabulary	4	1.1%	11 Unknown topic	3	0.8%
1 Faster reading	4	1.1%	1 Faster reading	0	0.0%	8 Interesting topic	1	0.3%
No answer provided	123	34.4%	No answer provided	150	41.9%	No answer provided	129	36.0%
Total	381		Total	374		Total	399	

It was felt that the effect of familiarity ratings could be further refined by collapsing *Easier to make connections* to *Easier to understand*, as explanations given by students appeared to point in the same direction. The top results for all three texts are presented in Table 5.26.

Table 5.26 Most common effects of familiarity

Star Struck	**15.1%** - It didn't show 12.6% - Easy to understand, Easy to make connections 9.8% - More difficult to understand
Bonds that Rock and Roll	**14.0%** - More difficult to understand 10.3% - Unknown/ difficult vocabulary 8.3% - Easy to understand, Easy to make connections 8.7% - It didn't show
Taking Stock	**18.7%** - Easy to understand, Easy to make connections 16.8% - More difficult to understand 9.8% - Unknown/ difficult vocabulary

Similar to analysing difficulty ratings and attributions, the next phase was doing bivariate analyses for each text by cross-tabulating frequencies of answers to open questions with frequencies of familiarity ratings. The objective was to find out whether there were any trends in answers among students who reported being more or less familiar with the topic. The aim was also to compare the results between the texts.

Bivariate analyses of the effects of familiarity for the three texts are presented in Tables 5.27, 5.28, and 5.29. It is noteworthy that a large number of students didn't answer the question (150 for BRR, 129 for TS, and 123 for SS), and students who did not provide any answer to this question perceived their topic familiarity as low or medium.

Cross-tabulation of familiarity ratings and attributions: "Star Struck"

Main observations as regards familiarity ratings and their effects for the "Star Struck" text presented in Table 5.27 are as follows:

- The first most frequent answer *It didn't show* was given by 54 students, and 36 among them (or 66.7%) rated their familiarity as very low (1–2), with a median of 2.00.

- The second most frequent answer *Easier to understand* was given by 38 students, with familiarity ratings relatively dispersed from 1–5, and a median of 4.00.
- The third most frequent answer *More difficult to understand* was given by 35 students, of whom 30 (or 85.7%) rated their familiarity as very low (1–2), with a median of 1.00.

Table 5.27 Bivariate analysis of the "Star Struck" text

In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	**How familiar** were you with the topic before you read the text?							**Total**	**Median**
	Star Struck								
	1	**2**	**3**	**4**	**5**	**6**	**7**		
10 It didn't show	13	23	9	7	1	1	0	54	2.00
	24.1%	42.6%	16.7%	13.0%	1.9%	1.9%	0.0%	100.0%	
3 Easier to understand	7	4	7	7	12	1	0	38	4.00
	18.4%	10.5%	18.4%	18.4%	31.6%	2.6%	0.0%	100.0%	
4 More difficult to understand	18	12	3	0	2	0	0	35	1.00
	51.4%	34.3%	8.6%	0.0%	5.7%	0.0%	0.0%	100.0%	
6 Unknown/difficult vocabulary	11	8	7	0	0	0	0	26	2.00
	42.3%	30.8%	26.9%	0.0%	0.0%	0.0%	0.0%	100.0%	
2 Slower reading	12	6	0	0	0	0	0	18	1.00
	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
13 Other	2	2	4	5	2	1	1	17	4.00
	11.8%	11.8%	23.5%	29.4%	11.8%	5.9%	5.9%	100.0%	
11 Unknown topic	6	4	3	2	1	0	0	16	2.00
	37.5%	25.0%	18.8%	12.5%	6.3%	0.0%	0.0%	100.0%	
8 Interesting topic	3	0	2	6	4	0	0	15	4.00
	20.0%	0.0%	13.3%	40.0%	26.7%	0.0%	0.0%	100.0%	
7 Known/easy vocabulary	2	1	2	3	5	0	0	13	4.00
	15.4%	7.7%	15.4%	23.1%	38.5%	0.0%	0.0%	100.0%	
12 Known topic	0	2	0	1	3	3	0	9	5.00
	0.0%	22.2%	0.0%	11.1%	33.3%	33.3%	0.0%	100.0%	
5 Easier to make connections	1	1	2	1	1	1	0	7	3.00
	14.3%	14.3%	28.6%	14.3%	14.3%	14.3%	0.0%	100.0%	
9 Uninteresting topic	2	2	1	0	0	0	0	5	2.00
	40.0%	40.0%	20.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
1 Faster reading	0	0	0	1	3	0	0	4	5.00
	0.0%	0.0%	0.0%	25.0%	75.0%	0.0%	0.0%	100.0%	
Total	**77**	**65**	**40**	**33**	**34**	**7**	**1**	**257**	**2.00**
	30.0%	**25.3%**	**15.6%**	**12.8%**	**13.2%**	**2.7%**	**0.4%**		

Cross-tabulation of familiarity ratings and attributions: "Bonds that Rock and Roll"

Main observations with respect to familiarity ratings and their effects for the "Bonds that Rock and Roll" text presented in Table 5.28 are as follows:

- The answer *More difficult to understand* came first of all codes for the BRR text and was provided by 49 out of 50 students (or 98.0%) who perceived their topic familiarity as low (1–3), with a median of 2.00.
- The second most frequent answer *Unknown/difficult vocabulary* was given by 37 students, 28 of whom (or 75.7%) perceived their topic familiarity as very low (1–2), with a median of 2.00.
- The answer *It didn't show* came third of all codes and was given by 31 students, and 20 of these (or 64.5%) rated their familiarity as lowest possible (1), with a median of 2.00. Students' replies and ratings in this category seem to suggest that even though students thought they were completely unfamiliar with the topic, they believed that this didn't affect their reading comprehension.

Table 5.28 Bivariate analysis of the "Bonds that Rock and Roll" text

In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	**How familiar** were you with the topic before you read the text?							**Total**	**Median**
	Bonds That Rock and Roll								
	1	**2**	**3**	**4**	**5**	**6**	**7**		
4 More difficult to understand	17	21	11	0	0	1	0	50	2.00
	34.0%	42.0%	22.0%	0.0%	0.0%	2.0%	0.0%	100.0%	
6 Unknown/difficult vocabulary	11	17	4	4	0	1	0	37	2.00
	29.7%	45.9%	10.8%	10.8%	0.0%	2.7%	0.0%	100.0%	
10 It didn't show	14	6	4	2	3	2	0	31	2.00
	45.2%	19.4%	12.9%	6.5%	9.7%	6.5%	0.0%	100.0%	
13 Other	5	6	5	4	5	2	0	27	3.00
	18.5%	22.2%	18.5%	14.8%	18.5%	7.4%	0.0%	100.0%	
3 Easier to understand	2	2	8	5	4	1	0	22	3.00
	9.1%	9.1%	36.4%	22.7%	18.2%	4.5%	0.0%	100.0%	
2 Slower reading	6	7	3	0	0	0	0	16	2.00
	37.5%	43.8%	18.8%	0.0%	0.0%	0.0%	0.0%	100.0%	
9 Uninteresting topic	3	3	4	0	0	0	0	10	2.00
	30.0%	30.0%	40.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
5 Easier to make connections	0	1	2	2	2	0	1	8	4.00
	0.0%	12.5%	25.0%	25.0%	25.0%	0.0%	12.5%	100.0%	
8 Interesting topic	3	1	1	2	0	0	0	7	2.00
	42.9%	14.3%	14.3%	28.6%	0.0%	0.0%	0.0%	100.0%	
11 Unknown topic	2	3	1	0	0	0	0	6	2.00
	33.3%	50.0%	16.7%	0.0%	0.0%	0.0%	0.0%	100.0%	
12 Known topic	1	2	1	0	1	1	0	6	2.50
	16.7%	33.3%	16.7%	0.0%	16.7%	16.7%	0.0%	100.0%	
7 Known/easy vocabulary	0	0	2	2	0	0	0	4	3.50
	0.0%	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%	100.0%	
Total	**64**	**69**	**46**	**21**	**15**	**8**	**1**	**224**	2.00
	28.6%	**30.8%**	**20.5%**	**9.4%**	**6.7%**	**3.6%**	**0.4%**	**100.0%**	

Cross-tabulation of familiarity ratings and attributions: "Taking Stock"

Main observations regarding the familiarity ratings and their effects for the "Taking Stock" text presented in Table 5.29 are as follows:

- The first most frequent answer *More difficult to understand* was given by 60 students, 48 of whom (or 80.0%) rated their familiarity as relatively low (1–3), with a median of 2.00.

- The second most frequent answer *Easier to understand* was given by 49 students, 42 of whom (or 85.7%) rated their topic familiarity as relatively low (1–3), with a median of 2.00.
- The third most frequent answer *Unknown/difficult vocabulary* was provided by 35 students, and 31 of these rated their familiarity as low or medium (1–3), with a median of 2.00.

Table 5.29 Bivariate analysis of the "Taking Stock" text

In what way did your **familiarity** with the topic or lack of it show in your reading comprehension?	**How familiar** were you with the topic before you read the text? Taking Stock							**Total**	**Median**
	1	**2**	**3**	**4**	**5**	**6**	**7**		
4 More difficult to understand	15	17	16	8	3	1	0	60	2.00
	25.0%	28.3%	26.7%	13.3%	5.0%	1.7%	0.0%	100.0%	
3 Easier to understand	11	20	11	3	2	1	1	49	2.00
	22.4%	40.8%	22.4%	6.1%	4.1%	2.0%	2.0%	100.0%	
6 Unknown/difficult vocabulary	10	10	11	3	0	0	1	35	2.00
	28.6%	28.6%	31.4%	8.6%	0.0%	0.0%	2.9%	100.0%	
10 It didn't show	8	4	2	2	5	2	0	23	2.00
	34.8%	17.4%	8.7%	8.7%	21.7%	8.7%	0.0%	100.0%	
7 Known/easy vocabulary	9	8	5	0	0	0	0	22	2.00
	40.9%	36.4%	22.7%	0.0%	0.0%	0.0%	0.0%	100.0%	
13 Other	5	7	3	2	1	1	0	19	2.00
	26.3%	36.8%	15.8%	10.5%	5.3%	5.3%	0.0%	100.0%	
5 Easier to make connections	4	6	5	1	2	0	0	18	2.00
	22.2%	33.3%	27.8%	5.6%	11.1%	0.0%	0.0%	100.0%	
12 Known topic	6	5	2	1	2	1	0	17	2.00
	35.3%	29.4%	11.8%	5.9%	11.8%	5.9%	0.0%	100.0%	
2 Slower reading	7	1	1	1	1	0	0	11	1.00
	63.6%	9.1%	9.1%	9.1%	9.1%	0.0%	0.0%	100.0%	
9 Uninteresting topic	3	0	2	1	2	0	0	8	3.00
	37.5%	0.0%	25.0%	12.5%	25.0%	0.0%	0.0%	100.0%	
1 Faster reading	1	2	1	0	0	0	0	4	2.00
	25.0%	50.0%	25.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
11 Unknown topic	3	0	0	0	0	0	0	3	1.00
	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
8 Interesting topic	0	1	0	0	0	0	0	1	2.00
	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
Total	**82**	**81**	**59**	**22**	**18**	**6**	**2**	**270**	**2.00**
	30.4%	**30.0%**	**21.9%**	**8.1%**	**6.7%**	**2.2%**	**0.7%**		

5.6 Bivariate correlation analyses of variables

To determine the strength of the relationships between the pairs of target variables (criterion and explanatory variables), bivariate analyses were carried out. As described in section 4.2.2, the current study used two operationalizations of background knowledge: a) a score on a test of background knowledge (BK), b) self-reports on familiarity with three text topics (SRF), expressed as ratings in the post-reading questionnaire. It was intended that combining the two different operationalizations of background knowledge would provide a better insight into the relationships between the studied variables. The following list of variables was analysed: the criterion variable Reading Comprehension and three explanatory variables Grammar Knowledge, Background Knowledge, and Self-Reported Familiarity. As outlined before, the data used for analyses referred to the three subject-specific texts (Star Struck (SS), Bonds that Rock and Roll (BRR), Taking stock (TS)) and were obtained from 358 students. It was anticipated that the results of correlation analyses (using a nonparametric Spearman's rho coefficient) should indicate to what degree independent variables co-vary with reading comprehension, as well as among themselves. Table 5.30 lists the results of bivariate correlation analyses and the following observations about the relationships among the measured variables can be made:

1. All three explanatory variables Background Knowledge, Grammar Knowledge, and Self-reported Familiarity correlated with the criterion variable, and the relationships were statistically significant (rho_{BK} = .471** (p <. 001), rho_{GK} = .412** ($p < .001$) and rho_{SRF} = .162** ($p = .002$)).
2. The strongest relationship was found between Reading Comprehension and Background Knowledge, with a correlation coefficient of .471** (a moderate association (see: Hopkins 2001, Cohen 1988, McGraw & Wong, 1992)) and significance at < .001 level.
3. Grammar Knowledge correlated with Reading Comprehension at .412** (a moderate association) and with significance at < .001 level. It appears that the strength is somewhat lower for the relationship between Reading Comprehension and Grammar Knowledge, yet the difference is relatively small.

4. The lowest statistically significant correlation was between Reading Comprehension and Self-reported Familiarity at .162** (a weak association) and with *p* = .002.
5. There were also statistically significant correlations among explanatory variables themselves, as Grammar Knowledge and Background Knowledge correlated at .321** (a weak to moderate association) with significance at < .001 level.
6. Further, Background Knowledge statistically significantly correlated with Self-reported Familiarity with *rho* = .196** (a weak association) and *p* < .001.
7. No correlation was found between Grammar Knowledge and Self-reported Familiarity (*rho* = .046, *p* = .390), indicating the absence of an association between the two variables.

Table 5.30 Bivariate correlations among Reading Comprehension, Grammar Knowledge, Background Knowledge, and Self-reported Familiarity (n=358)

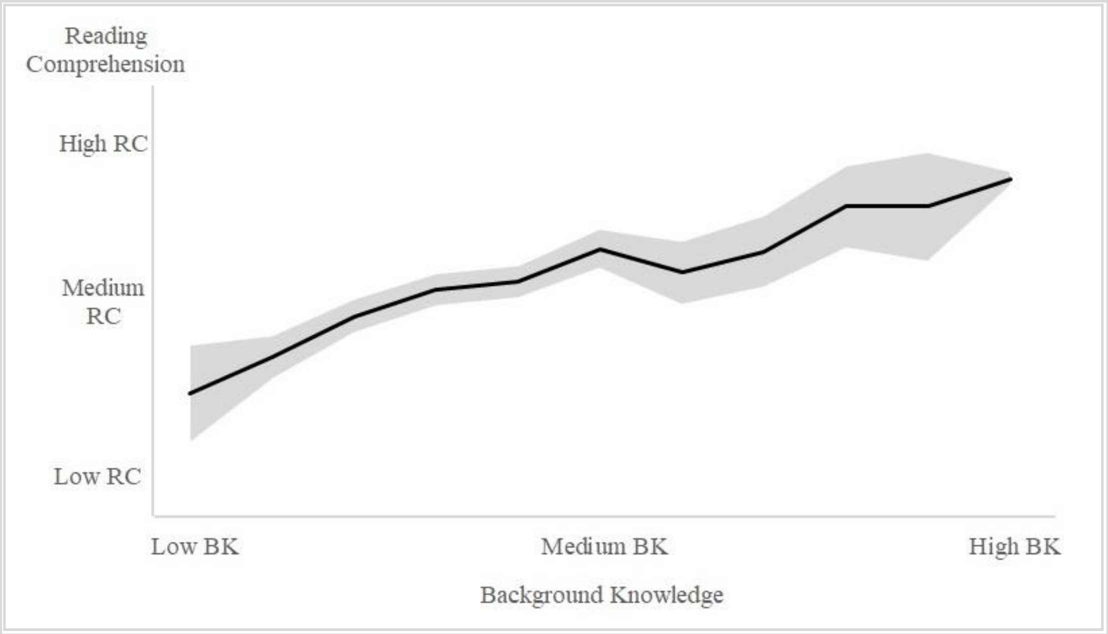
		Reading Comprehension	Grammar Knowledge	Background Knowledge	Self-reported familiarity
Reading Comprehension	Spearman's rho	1			
	p value				
	n				
Grammar Knowledge	Spearman's rho	.412**	1		
	p value	< .001			
	n	358			
Background Knowledge	Spearman's rho	.471**	.321**	1	
	p value	< .001	< .001		
	n	358	358		
Self-reported familiarity	Spearman's rho	.162**	.046	.196**	1
	p value	.002	.390	< .001	
	n	354	354	354	

**statistically significant at *p* < 0.01, *statistically significant at *p* < 0.05

Overall, bivariate analyses rendered a series of medium and low correlation coefficients indicating moderate and weak relationships between the studied variables. As correlations are measures of relatedness between variables, but they do not provide evidence of causality, further investigation was necessary to determine whether and to what extent there are effects of one variable on another. In order to do that multiple regression was used as a standard analysis in applied linguistics (Hatch & Lazaraton, 2000).

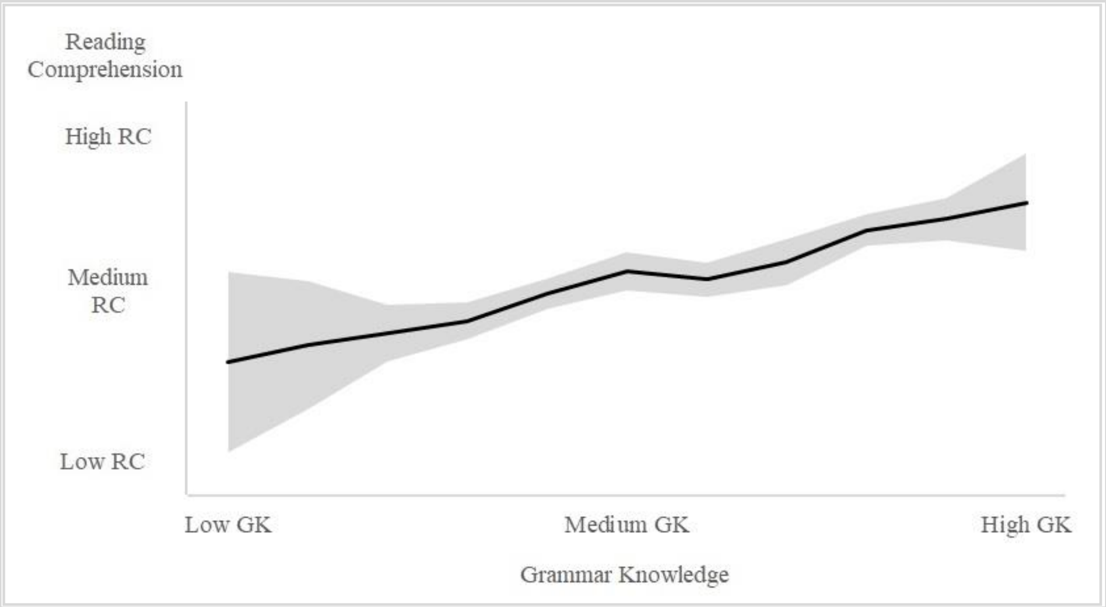
The data from the bivariate analyses were used to prepare a graphic representation of the relationships between the studied variables that are presented in Figures 5.5 and 5.6. They show two graphs summarizing the relationship between Background Knowledge and Reading Comprehension on the one hand, and Grammar Knowledge and Reading Comprehension on the other. The graph representing Background Knowledge was plotted according to students' scores on the background knowledge test, and their average scores on the reading comprehension test at a given level of background knowledge. Similarly, the Grammar Knowledge graph was plotted according to students' scores on the grammar knowledge test, and their average scores on the reading comprehension test at different levels of grammar knowledge. The y-axis shows Reading Comprehension, and the graph was plotted through the points that represent the Reading Comprehension score means at different levels of Background Knowledge and Grammar Knowledge.

Figure 5.5 Illustration of relationship between Background Knowledge (BK) and Reading Comprehension (RC) – average RC scores with 95% confidence intervals at different BK levels



The Background Knowledge graph in Figure 5.5 shows a relatively linear relationship between Background Knowledge and Reading Comprehension. In general, when Background Knowledge is higher, so is Reading Comprehension. As confidence intervals presented with grey colour are quite wide at very low BK and at high BK levels, any visual deviations from the linearity are a result of relatively small samples at certain BK levels.

Figure 5.6 Illustration of relationship between Grammar Knowledge (GK) and Reading Comprehension (RC) – average RC scores with 95% confidence intervals at different GK levels



The Grammar Knowledge graph presented in Figure 5.6 depicts the relationship between Grammar Knowledge and Reading Comprehension. The curve is showing a constant growth of Reading Comprehension at increased levels of Grammar Knowledge, and it shows a relationship which is almost linear. Again, confidence intervals are fairly wide, especially at low GK and high GK levels due to small samples at those levels.

5.7 Multiple regression analyses of constructs

Multiple regression was used to explore the relationship between the studied variables, and to test the contribution of explanatory variables (Background Knowledge, Grammar Knowledge, Self-reported Familiarity) to the performance on the criterion variable (Reading Comprehension). Multiple linear regression analysis was carried out with the aim to define which variables explain the most variance of the criterion variable, and may be therefore seen as the best predictors of success at reading comprehension of subject-specific texts. More specifically, the answers to the following research questions were sought:

RQ 1: Do Background Knowledge, Grammar Knowledge and Self-reported Familiarity affect Reading Comprehension of subject-specific texts?

RQ 2: To what extent do Background Knowledge, Grammar Knowledge and Self-reported Familiarity explain individual differences in Reading Comprehension of subject-specific texts?

Table 5.31 presents the results of multiple linear regression analyses. The *Adjusted R Square* is .295, suggesting that 29.5% of the variance of the criterion variable can be explained by independent variables of the regression model. While this is a good result, it also means that a substantial proportion of variance remains unexplained.

The beta coefficients show a statistically significant influence of Background Knowledge (β = .466, p < .001) and Grammar Knowledge (β = .552, p < .001) on Reading Comprehension. There is no association between Self-reported Familiarity and Reading Comprehension (β = .174, p = 0.124) in a multivariate setting, which differs from the bivariate, i.e. correlation analysis results (see Table 5.30). This shows the importance of testing the effect of predictor variables on Reading Comprehension with multivariate analyses.

Since I would also like to compare the magnitude of the effect of Background Knowledge and Grammar Knowledge on Reading Comprehension, and potential ranges of predictor variable values differ (GK: 0-30, BK: 0-42), I am presenting standardized beta coefficients in the last column of Table 5.31, as they can be directly compared. The standardized beta coefficient for Background Knowledge is $\beta = .343$ ($p < .001$), and for Grammar Knowledge $\beta = .312$ ($p < .001$). These results indicate that the contribution and predictive power of Background Knowledge could be slightly stronger than that of Grammar Knowledge, but the difference is very small.

Table 5.31 Multiple regression model, outcome variable: Reading Comprehension, predictors: Grammar Knowledge, Background Knowledge, Self-reported Familiarity (n=354)

	Coefficient	**Std. error (coefficient)**	***Standardized coefficient***
Grammar Knowledge	.552**	.083	*.312*
Background Knowledge	.466**	.066	*.343*
Self-Reported Familiarity	.174	.113	*.071*
Constant	18.036	1.793	
Adjusted R Square	*.295*		

*statistically significant at $p < 0.05$, **statistically significant at $p < 0.01$

5.8 Summary of bivariate and multivariate analyses and answers to research questions

This section addressed the research questions and reported the findings of additional statistical analyses, both bivariate (i.e. Kruskal-Wallis H test and Mann-Whitney U test) and multivariate (i.e. multiple linear regression analysis and multilevel mixed-effects generalized linear modelling).

5.8.1 Extending the analysis of the impact of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity on Reading Comprehension

First, correlation analyses pointed to a moderate relationship between Background Knowledge and Reading Comprehension, and to a slightly weaker moderate relationship between Grammar Knowledge and Reading Comprehension. The findings also suggested a moderate intercorrelation between Background Knowledge and

Grammar Knowledge, as well as a weak correlation between Self-reported Familiarity and Background Knowledge.

Second, after ascertaining correlations between variables, the next step was to explore the direction of the influences among variables by multiple linear regression and define the amount of total variability of Reading Comprehension scores that can be explained by Background Knowledge, Grammar Knowledge and Self-reported Familiarity. The regression model accounted for 29.5% of variance in the criterion variable that could be explained by predictor variables. Reading Comprehension of subject-specific texts could be predicted by Background Knowledge and to a slightly lesser extent by Grammar Knowledge, as contributions of both variables were statistically significant. On the other hand, even though correlational analyses indicated a relationship between Self-reported Familiarity and Reading Comprehension, this was not corroborated by the regression analysis, as there was no evidence of statistically significant influence in a multivariate context.

Third, as the regression model presented in Table 5.31 refers to the total sample and results for all three texts together, it was felt necessary to extend the research in two directions:

a) Comparison of the contribution of variables for each of the three texts separately: SS, BRR, TS.
b) Exploration of the impact of Background Knowledge and Grammar Knowledge and Self-reported Familiarity on Reading Comprehension if students are divided into subgroups according to their scores on the background knowledge and grammar knowledge tests.

Regression analysis was used for both purposes and the results are outlined in sections 5.8.2 and 5.8.3. The results will provide the required evidence to address the research questions.

5.8.2 The impact of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity on Reading Comprehension of individual texts

Having examined the regression data on the impact of variables referring to all texts together, it was felt necessary to see whether the results vary among the three texts and if so, to what extent. This section thus addresses the following research sub-question:

Sub-question: Does the contribution of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity to Reading Comprehension of subject-specific texts vary among the three texts?

To answer the sub-question, a new multilevel regression model (more specifically, multilevel mixed-effects generalized linear model) was constructed. The aim was to test the impact of three explanatory variables (Background Knowledge, Grammar Knowledge, and Self-reported Familiarity) on the criterion variable Reading Comprehension of a reading text (SS, BRR and TS, separately). The objective was also to check whether there are differences among texts.

After excluding units (i.e. students) with missing information for Self-reported Familiarity, I included Reading Comprehension scores (as the dependent variable) for the three texts and 356 students in the new multilevel regression model. Please note that the new dependent variable measures a relative score for each text (range 0-1 (or 0%-100%), see Table 5.4, column “Mean %”) and not a total score (which was calculated for three texts combined, see Table 5.3). Hence, beta coefficient values from the two regression models (from Tables 5.31 and 5.32) cannot and should not be directly compared.

In addition, to address the research questions, we expanded the range of predictors of individual text Reading Comprehension scores by adding the following independent variables: 1) texts, 2) text-GK interaction term, 3) text-BK interaction term, 4) self-reported familiarity with individual texts (e.g. familiarity with SS predicts RC SS in the model). Table 5.32 summarizes the results of the new regression model predominantly focusing on the impact of the three reading texts.

Table 5.32 Contribution of Background Knowledge, Grammar Knowledge and Self-reported Familiarity to outcome variable Reading Comprehension of Individual Texts, multilevel regression analysis (n=356)

Predictor variable		Coefficient	Std. error
Grammar Knowledge		0.007**	0.002
Background Knowledge		0.006**	0.002
Text	Taking Stock	0	
	Star Struck	-0.048	0.051
	Bonds That Rock and Roll	-0.357**	0.051
Text-GK interaction	GK with Taking Stock	0	
	GK with Star Struck	0.004	0.003
	GK with Bonds That Rock and Roll	0.010**	0.003
Text-BK interaction	BK with Taking Stock	0	
	BK with Star Struck	0.003	0.002
	BK with Bonds That Rock and Roll	0.008**	0.002
Self-reported Familiarity *(with individual texts)*[a]		0.009*	0.004
Constant		0.471	0.046

*statistically significant at $p < 0.05$, **statistically significant at $p < 0.01$

[a] self-reported familiarity measured with a horizontal rating scale is used as an interval (numeric) predictor

The results in Table 5.32 indicate that the level of influence of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity on Reading Comprehension of a particular text differs among the three texts. We can confirm a general influence of Grammar Knowledge and Background Knowledge on Reading Comprehension for all texts (β coefficients of .007 and .006 at $p < .001$). However, there are four notable differences among the texts and text-specific effects of Grammar Knowledge and Background Knowledge on Reading Comprehension:

1) "Bonds that Rock and Roll" text has a lower Reading Comprehension score than the other two texts, all else held constant (β = - .357 at $p < .001$)

2) Grammar Knowledge has a greater effect on "Bonds that Rock and Roll"-specific Reading Comprehension than on Reading Comprehension for the other two texts (β = .010 at $p < .001$).

3) Background Knowledge also has a greater effect on "Bonds that Rock and Roll"-specific Reading Comprehension score than on Reading Comprehension scores for the other two texts (β = .008 at p < .001).

4) Self-reported Familiarity with specific texts has a positive effect on Reading Comprehension scores for individual texts (β = .009 at p = .020)

Comparing the two multiple linear regression models (from Tables 5.31 and 5.32), we can observe the following differences in findings:

a) While Background Knowledge and Grammar Knowledge have a positive effect on Reading Comprehension in both models, the effect can be greater for specific texts – in this case, that was "Bonds that Rock and Roll", which stood out as the text with the lowest relative Reading Comprehension score, for which participants reported lower average level of familiarity, high level of difficulty, and a greater effect of low familiarity on Reading Comprehension (compared to SS and TS texts).
b) While Self-reported Familiarity for all three texts combined (see the regression model from Table 5.31) did not influence Reading Comprehension for all three texts combined, Self-reported Familiarity for individual texts (see the regression model from Table 5.32) had a positive effect on Reading Comprehension scores calculated for individual texts (e.g. SRF_SS → RC_SS).

To recap, the regression data exploring the variables with reference to individual texts revealed that both Background Knowledge and Self-reported Familiarity for individual texts have a statistically significant impact on Reading Comprehension for individual text. However, since the range of valid values was much wider for Background Knowledge (i.e. 0-30, see Table 5.2) than for Self-reported Familiarity (i.e. 1-7, see Tables 5.8, 5.10 and 5.12), and beta coefficients were comparable (β_{BK} = .006, β_{SRF} = .009), we can conclude that the impact of Background Knowledge on Reading Comprehension is greater than the impact of Self-reported Familiarity on Reading Comprehension for individual texts. It is notable that this is consistent with the correlation analysis results for the combined Reading Comprehension and Self-reported Familiarity scores (see Table 5.30).

5.8.3 Regression analysis: the impact of variables referring to subgroups of students

The relationship between the main variables was further tested by dividing students into groups according to their Background Knowledge and Grammar Knowledge, to observe the influence of the varying degrees of Background Knowledge or Grammar Knowledge on Reading Comprehension. This was done as there was evidence from previous studies that background knowledge can only be used at certain levels of grammar knowledge. Based on such findings, one- or two-threshold hypotheses were explored and posited in a number of studies (Hudson, 1982; Alderson & Urquhart, 1985/8; Hynd & Alverman, 1989; Hung, 1990; Clapham, 1996; Ridgway, 1997; Krekeler, 2006; Uso-Juan, 2006; Stevenson, Shoonen & de Glopper, 2007). In the literature, the lower threshold meant that students below a certain level of grammar knowledge could not use their background knowledge, whereas the upper threshold indicated that above a certain point of grammar knowledge students no longer relied on their background knowledge. To see whether such claims could be supported by the evidence obtained in the present study, the following two research questions were addressed:

RQ 3: Does the contribution of Background Knowledge to Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

RQ 4: To what extent does the contribution of Background Knowledge on Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

Hypothesis: Students can use their Background Knowledge only at a certain level of Grammar Knowledge

Regression was used to test whether and to what extent the influence of Background Knowledge on Reading Comprehension varies in different groups of students according to their Grammar Knowledge. As presented in Table 5.33 the whole sample was divided into three approximately equal groups according to students' Grammar Knowledge.

Table 5.33 Division into groups according to students' Grammar Knowledge

	Group 1 Low GK, scores 1-17	**Group 2** Medium GK, scores 18-21	**Group 3** High GK, Scores 22-30
n	120	128	110
GK score mean	14.74	19.55	24.45
GK score SD	2.21	1.16	2.09
BK score mean	6.73	8.63	11.25
BK score SD	4.63	5.16	6.12
RC score mean	30.83	34.38	37.85
RC score SD	7.74	7.12	6.46
SRF score mean *(all texts combined)*	8.17	7.90	8.50
SRF score SD *(all texts combined)*	2.86	3.15	3.27

The students with lower Grammar Knowledge are in Group 1 (n=120), those with medium Grammar Knowledge are in Group 2 (n=128), and students with the best Grammar Knowledge are in Group 3 (n=110). The table displays Grammar Knowledge score means for each group, which is the lowest in Group 1 (range 1-17, mean=14.74), medium in Group 2 (range 18-21, mean=19.55) and highest in Group 3 (range 22-30, mean=24.45). A similar trend can be observed in the Background Knowledge and Reading Comprehension mean scores: in Group 1 the Background Knowledge mean score is 6.73, in Group 2 it is 8.63 and in Group 3 it is 11.25. Reading Comprehension mean score is the lowest in Group 1 (30.83), slightly higher in Group 2 (34.38) and the highest in Group 3 (37.85). The comparison of the distribution of scores suggests that the groups show a common trend in all variables, which is consistent with the reported statistically significant associations between pairs of those three studied variables (see Table 5.30). On the other hand, the Self-reported Familiarity mean score is the lowest in Group 2 (7.90), and the highest in Group 3 (8.50).

All students were also divided into three approximately equal groups according to their Background Knowledge scores as shown in Table 5.34. The students with lower

Background Knowledge are in Group 1 (n=136), those with medium Background Knowledge are in Group 2 (n=109), and students with the best Background Knowledge are in Group 3 (n=113). The aim was to create three groups with three levels of Background Knowledge that were of similar size and as homogeneous as possible. This was expected to allow a comparison between the groups with the highest and the lowest Background Knowledge and Grammar Knowledge, such as "Low BK-High GK" group and "High BK-Low GK" group. The comparison of "extreme" groups is crucial when testing the compensation effect.

Table 5.34 Division into groups according to students' Background Knowledge

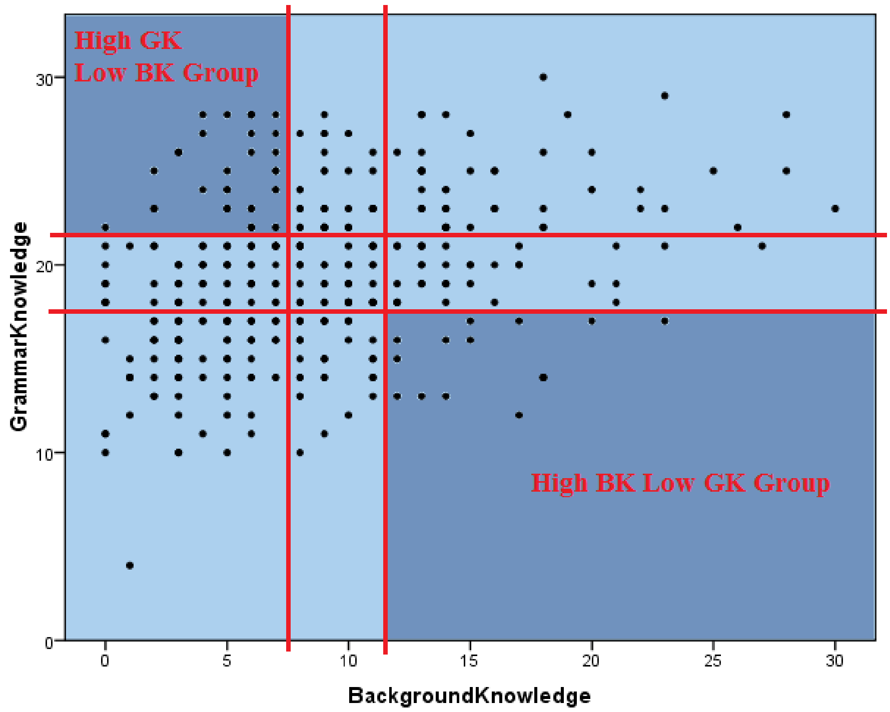
	Group 1 Low BK, scores 1-6	**Group 2** Medium BK, scores 7-11	**Group 3** High BK, scores 12-30
n	136	109	113
BK score mean	3.68	8.37	15.36
BK score SD	1.91	1.05	4.36
GK score mean	18.06	19.68	20.88
GK score SD	4.43	3.93	4.02
RC score mean	30.30	35.17	38.12
RC score SD	7.65	6.83	6.05
SRF score mean *(all texts combined)*	7.65	7.84	9.11
SRF score SD *(all texts combined)*	2.71	3.04	3.38

Table 5.34 shows the different Background Knowledge score ranges and means for the groups: Group 1 (range 1-6, mean=3.68), Group 2 (range 7-11, mean=8.37) and Group 3 (range 12-30, mean=15.36). Similarly, there are differences in values for Grammar Knowledge, Reading Comprehension, and Self-reported Familiarity between Groups 1, 2 and 3 (18.06, 19.68 and 20.88 for GK; 30.30, 35.17, and 38.12 for RC; 7.65, 7.84 and 9.11 for SRF), which even at face value show a relationship between the variables (which is consistent with the results presented in Table 5.30).

Figure 5.7 shows the distribution of students into groups according to both their scores for Grammar Knowledge and Background Knowledge. It presents the distribution of students into groups according to their scores on the tests of Grammar Knowledge and

Background Knowledge. Nine groups were formed: three groups according to Grammar Knowledge scores, and three groups according to Background Knowledge scores (3x3). The divisions were made in such a way that there would be a sufficient number of students even in the most "extreme" groups (e.g. "High BK-Low GK" group and "High GK-Low BK" group). This meant that for instance "Low GK-High BK" group included those students who scored more than 11 points on the background knowledge test, and less than 18 points on the grammar knowledge test. On the other hand, High GK-Low BK group included those students who scored 7 points or less on the background knowledge test, and more than 21 points on the grammar knowledge test.

Figure 5.7 Visual division into groups according to Grammar Knowledge and Background Knowledge



Although these divisions into subgroups were used for comparisons among groups, it is obvious that the differences in scores (either Grammar Knowledge, Background Knowledge or Reading Comprehension) were relatively small and because of that, conclusions need to be interpreted with caution. First, to address RQs 3 and 4, another multiple linear regression model needed to be constructed. It resembles the multiple regression model presented in Table 5.31, as it consists of the same outcome variable (i.e. Reading Comprehension score for all texts combined) and three matching independent variables (Grammar Knowledge, Background Knowledge, and Self-reported Familiarity scores). However, to provide answers to the research questions, interaction terms “BK group-GK score” and “GK group-BK score” had to be added to determine at what Grammar Knowledge level(s) students can efficiently use their Background Knowledge (and vice versa). The groupings presented in Tables 5.33 and 5.34 were used in those two interaction terms.

Table 5.35 Multiple regression model, outcome variable: Reading Comprehension, predictors: Grammar Knowledge, Background Knowledge, Self-reported Familiarity, “BK group-GK score” and “GK group-BK score” interaction terms (n=354)

Predictors		**Coefficient**	**Std. error**
Grammar Knowledge		0.785**	0.123
Background Knowledge		0.470**	0.133
BK groups – GK score interaction	GK at Medium BK	0	
	GK at Low BK	-0.079	0.051
	GK at High BK	-0.034	0.058
GK groups – BK score interaction	BK at Medium GK	0	
	BK at Low GK	0.098	0.108
	BK at High GK	-0.172*	0.084
Self-Reported Familiarity (with all texts combined)		0.201	0.113
Constant		14.397**	2.603
Adjusted R Square		*0.305*	

*statistically significant at $p < 0.05$, **statistically significant at $p < 0.01$

The results from Table 5.35 show that the model explains a similar amount of total variability (30.5%) of Reading Comprehension scores as the initial regression model presented in Table 5.31 (29.5%) and confirm the influence of Background Knowledge (β = .470 at p < .001) and Grammar Knowledge (β = .785 at p < .001) on Reading Comprehension, but not of the combined Self-reported Familiarity (β = .201 at p = .075).

There are additional results presented in Table 5.35 which can be used to address Research Questions 3 and 4:

1) On using Background Knowledge at different levels of Grammar Knowledge: comparing the use of Background Knowledge at Medium GK to that at Low or High GK levels, a negative effect can be reported for High GK level (β = - .172 at p = .041). This means that at High GK level, Background Knowledge is used slightly less effectively than at Medium GK and especially Low GK level.
2) On using Grammar Knowledge at different Background Knowledge levels: comparing the use of Grammar Knowledge at Medium BK to that at Low or High BK levels, no statistically significant effects can be confirmed. This suggests that students can effectively use their grammar knowledge to read L2 texts irrespectively of their background knowledge.

However, since the "general" Background Knowledge beta coefficient value equals to 0.470, and it is much higher absolutely than the beta coefficient for "BK at High GK group" (β = - .172), we have to reject the hypothesis that students can use their Background Knowledge only at a certain level of Grammar Knowledge. Ultimately, the evidence from Table 5.35 suggests that students can use their Background Knowledge at all levels of Grammar Knowledge, but the role of Background Knowledge is stronger at lower levels of Grammar Knowledge, possibly implying compensation.

5.8.4 The compensatory effect of Background Knowledge and Grammar Knowledge

A number of studies have examined the compensatory effect in reading comprehension (Stanovich, 1980, 2000; Alderson & Urquhart, 1985; Koh, 1985; Goldman & Duran, 1988, Chen & Donin, 1997; Ridgway, 1997; Yamashita, 2002; Uso-Juan, 2006; Bernhardt, 2011) They posited compensatory models, suggesting that students could compensate for lower grammar knowledge with higher background knowledge and vice versa. To see whether such claims can be substantiated by the results obtained in the present study, the following questions were addressed:

RQ 5: Can students with better Background Knowledge compensate for lower Grammar Knowledge in Reading Comprehension of subject-specific texts?

RQ 6: Can students with better Grammar Knowledge compensate for lower Background Knowledge in Reading Comprehension of subject-specific texts?

Hypothesis: Better Grammar Knowledge allows students to compensate for lower Background Knowledge, and vice versa.

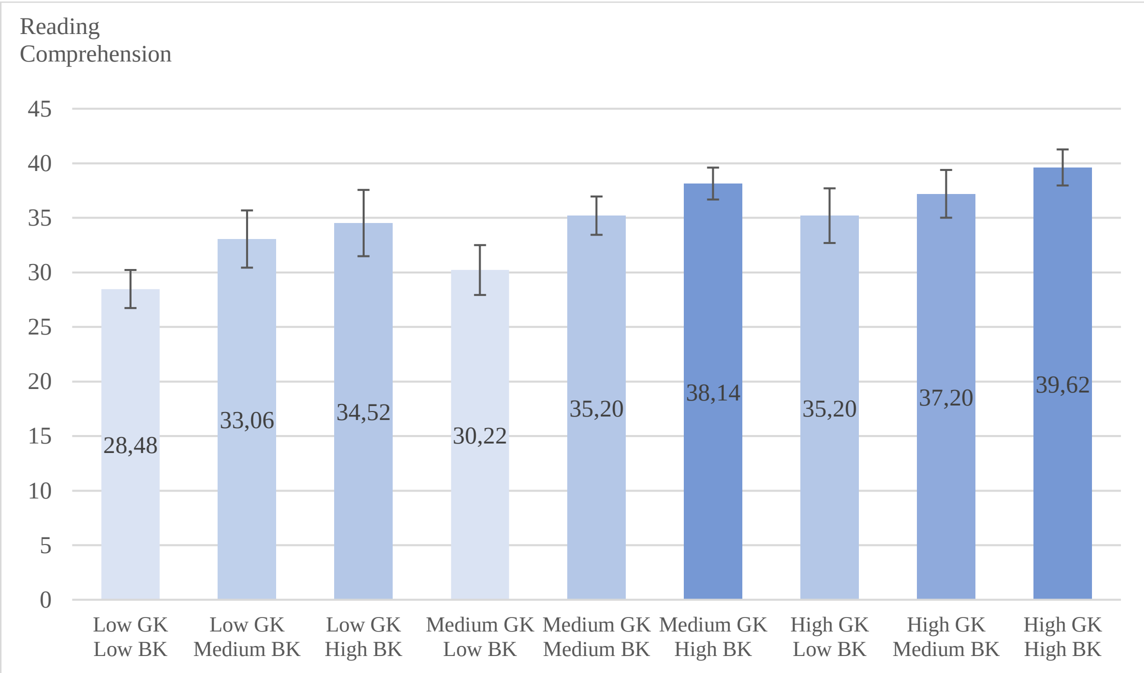
In order to answer RQs 5 and 6, students were divided into nine groups according to their Grammar Knowledge and Background Knowledge as presented in Tables 5.33 and 5.34 and Figure 5.7. These group divisions were then used to form the following groups:

1) "Low GK - Low BK" group *(n* = 65)
2) "Low GK - Medium BK" group *(n* = 34)
3) "Low GK - High BK" group *(n* = 21)
4) "Medium GK - Low BK" group *(n* = 46)
5) "Medium GK - Medium BK" group *(n* = 40)
6) "Medium GK - High BK" group *(n* = 42)
7) "High GK - Low BK" group *(n* = 25)
8) "High GK - Medium BK" group *(n* = 35)
9) "High GK - High BK" group *(n* = 50)

The two main groups of interest, i.e. "Low GK - High BK" and "High GK - Low BK" are the smallest in size, which is consistent with the positive association of Grammar Knowledge and Background Knowledge (see Table 5.30) – in practice this means that students with more Grammar Knowledge are more likely to have more Background Knowledge as well, and vice versa.

Figure 5.8 presents the mean Reading Comprehension scores across the nine BK-GK groups. These results confirm that average Reading Comprehension scores increase with both increased Grammar Knowledge levels and increased Background Knowledge levels, and even more if both Background Knowledge and Grammar Knowledge levels increase at the same time. For example, the average Reading Comprehension score for the "Low GK - Low BK" group is 28.48 (95% confidence interval (CI) [26.73, 30.22]), which increases to 35.20 (95% CI [32.70, 37.70]) for the "Low GK - High BK" group or to 34.52 (95% CI [31.49, 37.56]) for the "High GK - Low BK" group. This indicates that better Grammar Knowledge allows students to compensate to some extent for lower Background Knowledge, and that better Background Knowledge allows other students to compensate to some extent for lower Grammar Knowledge. The compensatory effect can be confirmed for both Background Knowledge and Grammar Knowledge and it appears comparable.

Figure 5.8 Mean Reading Comprehension of groups with different levels of Background Knowledge and Grammar Knowledge (n=358)



To test the hypotheses, statistical testing needs to be carried out. As the subsample sizes are fairly small (between 21 and 65 cases), it is prudent to combine some of them to have sufficient statistical power for statistical testing to detect any differences in the group means, while having the hypothesis in mind. The following GK-BK groups will be compared with Kruskal-Wallis H test:

1) “Compensating” groups *(i.e. “Low GK - High BK” + “High GK - Low BK” groups, n = 46)*
2) “Low GK or Low BK, no High” groups *(i.e. “Low GK – Low BK” + “Low GK - Medium BK” + “Medium GK - Low BK”, n =145)*
3) “High GK or High BK, no Low” groups *(i.e. “Medium GK - High BK” + “High GK - Medium BK” + “High GK - High BK”, n =127)*

All groups except for “Medium GK - Medium BK” group (n=40), which is sitting in the middle (also in terms of the mean Reading Comprehension score, see Figure 5.8), were classified in one of the three groups for multiple group comparison. The Kruskal-Wallis H test results are presented in Table 5.36. This nonparametric test is used (instead of ANOVA) since Reading Comprehension score is not a normally distributed numeric variable.

Table 5.36 Kruskal-Wallis H test results for mean Reading Comprehension scores, comparison of different GK-BK groups (n=318)

Categorical variable value*	n	Mean rank	Kruskall Wallist H test statistic	p value
“Compensating groups” (Low GK-High BK, High BK-Low GK)	46	165.72^{BC}	78.067 (df = 2)	< 0.001
“Low GK or Low BK, no High” groups	145	112.47^{AC}		
“High GK or High BK, no Low” groups	127	210.95^{AB}		

$^{A\,B\,C}$ = indicating statistically significant differences in RC scores between the groups at p<0.05 level (^{A}=“Compensating groups”, ^{B}= “Low GK or Low BK, no High” groups, ^{C}=“High GK or High BK, no Low” groups), * “Medium GK-Medium BK” group was excluded from this analysis

Kruskal-Wallis H test results show that there is a statistically significant difference in Reading Comprehension score between the analysed GK-BK groups ($p < 0.001$). The “Compensating groups” are significantly different from the other two groups, which is confirmed by pairwise comparisons. Students from “Compensating groups” on average achieved higher Reading Comprehension scores (mean rank = 165.72) than students who did not have high Background Knowledge scores or a high Grammar Knowledge scores (mean rank = 112.47). This means that we can report compensatory effects of both high Background Knowledge (in the case of low Grammar Knowledge) and high Grammar Knowledge (in the case of low Background Knowledge), and we can confirm the tested hypothesis. However, the evidence suggests that students from the “Compensating groups” cannot, on average, achieve the same Reading Comprehension scores as students from “High GK or High BK, no Low” groups (mean rank = 210.95). It can be argued that the compensatory effect is confirmed and significant.

5.9 Univariate and bivariate analysis of Difficulty, Self-reported Familiarity and Interest from the Post-Reading Questionnaire

Section 5.5 presented the post-reading questionnaire: its purpose, format, and descriptive statistics of the data referring to each of the three texts used for reading comprehension: Star Struck (SS), Bonds that Rock and Roll (BRR), and Taking Stock (TS). The aim of this section is to present the results of the bivariate analysis that addressed the question of whether there is a relationship between the criterion variable Reading Comprehension and explanatory variables: a) students' ratings of the difficulty of texts ("Difficulty All Texts", b) their reports of their familiarity with the text topics ("Self-Reported Familiarity All Texts", presented in section 5.5.5), and c) their interest in the texts ("Interest All Texts"). The reliability analysis of these instruments is presented in Chapter 4.

First, the descriptive statistics are presented in Table 5.37, and second, the correlations between Reading Comprehension, Difficulty, Self-Reported Familiarity, and Interest pertaining to individual texts, as well as to the three texts together, are presented in Table 5.38. Students' difficulty ratings of individual texts were collapsed into the construct "Difficulty All Texts"; this construct has a normal distribution, with a range of values from 4 to 21, a mean of 12.85, and a standard deviation of 3.076. Similarly, students' familiarity ratings for individual texts were used to form the construct "Familiarity All Texts" and its distribution is fairly normal, with a slight asymmetry on the right. It covers a range of values between 3 and 18, with a mean of 8.18 and a standard deviation of 3.097. Finally, the same procedure was done to obtain the construct "Interest All Texts", which is also normally distributed and has a range of values between 3 and 21, a mean of 11.50, and a standard deviation of 3.271. See Table 5.37 for further information.

Table 5.37 Descriptive statistics: Difficulty, Self-Reported Familiarity, Interest

	N	Min.	Max.	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
							Statistic	Std. Error	Statistic	Std. Error
Difficulty All Texts	353	4	21	12.85	3.076	9.459	-.116	.130	.092	.259
Familiarity All Texts	354	3	18	8.18	3.097	9.590	.341	.130	-.241	.259
Interest All Texts	313	3	21	11.50	3.271	10.700	-.025	.138	-.217	.275

Furthermore, the correlations between Difficulty, Self-Reported Familiarity, and Interest on the one hand, and Reading Comprehension on the other, for the three texts (SS, BRR, TS) are presented in Table 5.38. The table shows Spearman's rho coefficients and the corresponding statistical significance values of the relationships among the variables.

Table 5.38 Correlations: Difficulty, Self-reported Familiarity, Interest referring to the three texts (SS, BRR, TS) and Reading Comprehension scores

		Reading Comprehension
How **difficult** did you find the text? **SS**	Spearman's rho	-.276**
	p value	<0.001
	n	358
How **familiar** were you with the topic before you read the text? **SS**	Spearman's rho	.042
	p value	.426
	n	356
How **interesting** did you find the text? **SS**	Spearman's rho	.167**
	p value	.003
	n	321
How **difficult** did you find the text? **BRR**	Spearman's rho	-.195**
	p value	< 0.001
	n	355
How **familiar** were you with the topic before you read the text? **BRR**	Spearman's rho	.047
	p value	.379
	n	356
How **interesting** did you find the text? **BRR**	Spearman's rho	.172**
	p value	.002
	n	318
How **difficult** did you find the text? **TS**	Spearman's rho	-.209**
	p value	< 0.001
	n	357
How **familiar** were you with the topic before you read the text? **TS**	Spearman's rho	.253**
	p value	< 0.001
	n	357
How **interesting** did you find the text? **TS**	Spearman's rho	.082
	p value	.146
	n	318

**statistically significant at $p < 0.01$, *statistically significant at $p < 0.05$

The following observations can be made:

a) All three coefficients referring to the Difficulty of the three texts show a statistically significant but small negative correlation with Reading Comprehension, measured with Spearman's rho (*rho* = − .276 for SS, − .195 for BRR, and − .209 for TS, $p < .001$ for SS, BRR, and TS). The highest negative coefficient between the reported text Difficulty and Reading Comprehension is that of SS, and the lowest for BRR. This means that students scored better on Reading Comprehension of those texts which they found to be easier (SS), and they were somewhat right in their assessment of difficulty.

b) As for Self-reported Familiarity, only the correlation coefficient for the TS text is statistically significant (*rho* = .252, $p < .001$). No correlation was found between Self-reported Familiarity and Reading Comprehension for the other two texts. This means that only students who reported greater familiarity with the topic of the TS text also performed better on Reading Comprehension (for all three texts combined).

c) As regards the relation between students' assessment of how interesting they found the texts (Interest) and Reading Comprehension, there are two statistically significant correlations. They refer to the SS text (*rho* = .167, p = .003 and the BRR text (*rho* = .172, p = .002). The students who found the SS and BRR text more interesting also performed better on Reading Comprehension (for all three texts combined).

Having considered the correlations between the variables referring to individual texts, we can now move on to the aggregate correlations between the same variables for all the three texts together that are listed in Table 5.39.

Table 5.39 Correlations for all texts combined: Difficulty, Self-reported Familiarity, Interest, and Reading Comprehension

		Reading Comprehension All Texts	Self-reported Familiarity All Texts	Difficulty All Texts	Interest All Texts
Reading Comprehension All Texts	Spearman's rho	1			
	p value				
	n				
Self-reported Familiarity All Texts	Spearman's rho	.162**	1		
	p value	.002			
	n	354			
Difficulty All Texts	Spearman's rho	- .308**	- .347**	1	
	p value	< 0.001	< 0.001		
	n	353	350		
Interest All Texts	Spearman's rho	.192**	.405**	- .423**	1
	p value	.001	< 0.001	< 0.001	
	n	313	310	311	

**statistically significant at $p < 0.01$, *statistically significant at $p < 0.05$

The results reveal that there are statistically significant correlations between all variables. There is a weak and statistically significant correlation between "Self-reported Familiarity All Texts" and "Reading Comprehension All Texts" (*rho* = .162, *p* = .002). "Difficulty All Texts" and "Reading Comprehension All Texts" have a relatively small negative correlation (*rho* = - .308, *p* < .001), and a weak statistically significant correlation between "Interest All Texts" and RC (*rho* = .192, *p* = .001).

These results show that readers who found the texts to be more difficult had lower Reading comprehension scores, thus implying that they were right in their assessment of text difficulty. On the other hand, the readers who found the texts more interesting or reported to be more familiar with the text topics performed slightly better on the reading comprehension test. This suggests a positive relationship between topic interest, topic familiarity, and effective reading comprehension.

5.10 Reading habits and other variables

The post-reading questionnaire contained two questions about students' reading habits. They enquired about how much they generally read in English, or how much coursework reading in English they do. Correlation analysis was conducted to identify variables that reading habits are associated with.

5.10.1 Reading habits in relation to Grammar Knowledge, Background Knowledge and Reading Comprehension

Table 5.40 presents correlation analyses between both reading variables ("General Reading in English", "Course-related Reading in English") and the main variables: Grammar Knowledge, Background Knowledge, and Reading Comprehension. The following trends can be observed:

1. "General Reading in English" correlates weakly but significantly with "Course-related reading" (*rho* = .129, p = .015).
2. "General Reading in English" also correlates moderately and significantly with all other variables, most strongly with Grammar Knowledge (*rho* = .386, p < .001) and Reading Comprehension (*rho* = .355, p < .001). It also correlates with Background Knowledge (*rho* = .266, p < .001).
3. "Course-related Reading in English" does not correlate significantly with Grammar Knowledge and Reading Comprehension, just weakly with "General Reading in English", as noted before, and with Background Knowledge (*rho* = .141, p = .008).

Table 5.40 Correlations: Reading habits and main variables

		General Reading in English	Course-related Reading in English
General Reading in English	Spearman's rho	1	.129*
	p value		.015
	n		353
Course-related Reading in English	Spearman's rho	.129*	1
	p value	.015	
	n	353	
Grammar Knowledge	Spearman's rho	.386**	- .070
	p value	< 0.001	.190
	n	353	353
Background Knowledge	Spearman's rho	.266**	.141**
	p value	< 0.001	.008
	n	353	353
Reading Comprehension	Spearman's rho	.355**	.021
	p value	< 0.001	.690
	n	353	353

**statistically significant at p < 0.01, *statistically significant at p < 0.05

5.10.2 Reading habits and gender

The post-reading questionnaire data also provided the information about the gender of test-takers, and their a) "General Reading in English", and b) "Course-related Reading in English". Table 5.41 presents the results of the independent samples test that was run to see whether there are significant differences in "General Reading in English" among males and females. Nonparametric Mann-Whitney U test was used for statistical testing, as the reading habits were measured on an ordinal scale.

Table 5.41 Gender and reading habits ("General Reading in English" and "Course-related Reading in English"): Mann-Whitney U test

Scale variable	Categorical variable value	n	Mean rank	Mann-Whitney U	*p* value
General Reading in English	Female	213	148.83	8,910.50	< .001
	Male	128	207.89		
Course-related Reading in English	Female	213	179.23	15,386.00	0.042
	Male	128	157.30		

The results presented in Table 5.41 show statistically significant differences (at $p < .05$ level) between female and male students in both "General Reading in English" (mean rank, females = 148.83, males = 207.89, $p < .001$) and "Course-related Reading in English" (mean ranks, females = 179.23, males = 157.30, $p = .042$). The difference between males and females is bigger for "General Reading in English" (males read more) than for "Course-related Reading in English" (females read slightly more).

5.11 Chapter summary and conclusions

This chapter investigated six research questions together with sub-questions and hypotheses by presenting the findings and analyses that were used in the testing study. It aimed to show the relationships between the criterion variable (Reading Comprehension) and explanatory variables (Background Knowledge, Grammar Knowledge, and Self-reported Familiarity). The analyses used ranged from descriptive statistics to nonparametric bivariate correlation analyses, nonparametric bivariate tests, and multivariate analyses, including multiple linear regression. The chapter also presents the data obtained from analysing the post-reading questionnaire, which may shed additional light on the results.

CHAPTER 6
RESULTS: THE THINK-ALOUD STUDY

6.1 Chapter aim and overview

This chapter presents the analysis and results of the think-aloud (TA) study that involved 24 out of 382 students that took part in the study at the School of Business and Economics, University of Ljubljana. It looks into the relationship between background knowledge and reading comprehension, as shown by the qualitative analyses of the think-aloud data. Chapter 6 begins with a description of the think-aloud coding procedure and coding scheme used in the coding and analysis of the think-aloud protocols (i.e. TA1-TA24) obtained when students read three subject-specific texts: Star Struck (SS), Bonds that Rock and Roll (BRR), Taking Stock (TS). It continues with a summary of the think-aloud coding frequencies. This is followed by a comparison of processing patterns used by readers with high or low background knowledge and their impact on reading comprehension. Finally, answer to research question 7 is proposed.

6.2 The think-aloud coding procedure

The think-aloud protocols were transcribed verbatim and were coded by two coders: the financial expert involved in the study and myself. The coding scheme was devised directly from the think-aloud protocols and was data-driven. However, in order to make informed choices, I studied several coding taxonomies used in think-aloud research studies (Haarstrup, 1991; Crain Thoreson et al., 1997; Yamashita, 2002; Nassaji, 2003; Bråten & Strømsø, 2003). Pressley and Afflerbach’s (1995) taxonomy of categories of what readers do when reading is particularly comprehensive and informative. Their list of categories of reading strategies was used as a reference point, while the coding scheme in this study was derived directly from the data at hand in the think-aloud protocols.

The think-aloud coding scheme was refined after a pilot and subsequent discussion with the finance expert. I coded all think-aloud protocols with the modified categories and asked the finance expert to code 10% of protocols. The coded protocols were compared to test reliability of the think-aloud coding scheme. All differences in coding were discussed case by case in three joint sessions and a 100% agreement on the coding decisions was reached. After the coding and validation phase, I counted the frequencies of coding categories and summarized them in several tables. A comparison was made between the groups of readers that have high and low background knowledge in order to see whether different patterns and strategies were used by the two groups and how effective they were.

6.3 The think-aloud coding scheme

The following 17 categories were used in the coding of think-aloud protocols. Each category is described and examples from different think-aloud protocols are given for illustration. The readers thought out loud in Slovene, so parts of their think-aloud protocols given as examples were translated into English and are given in italics after the relevant part of the reading text.

Code 1: Correct paraphrase (CP)

This coding category refers to instances when readers correctly understand a phrase, clause, sentence, or section of the text. They paraphrase it, use synonyms or translate it correctly. Readers activate the correct linguistic and background knowledge from their long-term memory.

> Text: … the first movie to issue shares which will be tradeable …
> Think-aloud: *It will be the first film 'to issue shares,' they will launch shares that will be 'tradeable,' yeah, they will be listed on the market.* (TA4, BRR)

Code 2: Approximate paraphrase (AP)

Readers fail to grasp the full meaning of a phrase, clause, sentence, or section of the text, but they show partial understanding. This suggests that when reading, they retrieve some relevant linguistic or background knowledge from their long-term memory, yet fail to fully comprehend the phrase, clause, sentence, or section of the text.

> Text: Equity based pay …
> Think-aloud: *Pay which is based on the ownership of managers.* (TA2, TS)

Code 3: Wrong paraphrase (WP)

Readers produce a wrong paraphrase of the phrase, clause, sentence, or section of the text when thinking aloud. This category could also be labelled *wrong schema activation* or *distortion* because readers come up with a paraphrase that is based on either their lack of background knowledge or incorrect background knowledge.

> Text: From cinema to parking fines, almost everything is being securitized.
> Think-aloud: *From cinema, aaaam, tickets to parking, almost all things are insured*. (TA4, BRR)

Code 4: Correct translation (CT)

Readers focus only on one word and give the correct translation of the word into Slovene.

> Text: Default on payment.
> Think-aloud: *Not paying*. (TA23, BRR)

Code 5: Approximate translation (AT)

Readers focus only on one word and provide a partially correct translation of the word into Slovene.

> Text: Stock options.
> Think-aloud: *Bonds.* (TA12, TS)

In this case, the reader appears to know that ‘bonds’ and ‘stocks’ are both types of securities. However, because bonds and stocks are securities with distinctly different properties, the reader’s translation is coded as approximate.

Code 6: Wrong translation (WT)

Readers focus on one word and translate it incorrectly. This could be related to readers' lack of either linguistic or background knowledge. In other words, readers may either not be familiar with a particular word or the concept it stands for.

> Text: Default.
> Think-aloud: *Delay.* (TA17, BRR)

Code 7: Elaboration (ELAB)

This category covers instances when readers build connections between ideas in the text and their personal experience and knowledge by providing an additional explanation of what they have read or by giving examples, associations, and analogies in reaction to the text.

> Text: There may be moral issues of different sort as well. "What happens if the star is caught molesting kids?" asks Paul Taylor, of Duff & Phelps, a bond-rating agency.
>
> Think-aloud: *Moral issues. I'm thinking about Slovenia, it is the same here – a firm sponsors a celebrity, and then he does something stupid and tarnishes the image of the company. This is something pretty similar.* (TA8, BRR)

Code 8: Evaluation (EVAL)

This category is coded if readers focus on the worth of what has been processed by offering their opinions, feelings, approval, disapproval or criticism. Readers react evaluatively and make a value judgement about the content of the text or its style.

> Text: Investors got an 8% yield on bonds rated on investment grade.
> Think-aloud: *Investors got 8% interest on these papers. That's not at all bad.* (TA13, SS)

Code 9: Guessing (GUES)

Readers assign the meaning of the information in the text randomly and do not show the reasoning or logic behind their choice. According to Comer (2012), this category could also be described as "unexplained guessing" and is a differentiating characteristic of guessing vs inferencing. Whereas guessing is random and unexplained by the reader, inferencing is reasoned and explained (Tavakoli & Hyat, 2011).

> Text: We cannot know if boards expect or condone these sales.
> Think-aloud: *Well, they expect or find this normal. Condone - keep, probably*. (TA10, TS)

Code 10: Inferencing (INFR)

The category is coded when readers relate information encountered in the text to their knowledge, both linguistic and content. The interpretation they come up with is a result of readers' reasoning and conclusions about the text. This category was defined by Haarstrup (1991, p. 13) as "making informed guesses."

> Text: Mortgage payment …
> Think-aloud: *Mortgage is something to do with loans. You borrow something. So they were more focused on borrowing than actually making money*. (TA4, BRR)

Code 11: Repair (REP)

Readers deliberately return to a particular word, phrase, clause, or sentence in the text in order to restate their think-aloud. With hindsight, they attempt to repair their understanding, and they may be successful or not.

> <u>Text</u>: Prior ownership …
> <u>Think-aloud</u>: *It is about ownership. No, no, it is not present ownership but ownership in the past.* (TA2, TS)

Code 12: Signalling unknown word (SUW)

Readers explicitly point out that they do not understand a particular word. They may also overtly state that the word is new to them or is part of new or unfamiliar information in the text.

> Text: … ultimately, returns will depend on sales of DVDs, calendars and other paraphernalia.
> Think-aloud: *I don't know what 'paraphernalia'is.* (TA3, SS)

Code 13: Signalling lack of understanding (SLU)

Readers say that they do not understand a phrase, sentence, or section of the text because it is either new or difficult in some way. They may refer to problems in comprehension, for instance the fact that they may understand individual words but fail to grasp the meaning of a sentence.

> Text: There is barely a cash flow anywhere, it seems, that cannot be reassembled into a bond-like security that the most conservative of investors might buy.
> Think-aloud: *There is no money, it cannot be 'reassembled,' collected 'into a bond-like security.' These most conservative investors would buy something. I understand every word but I don't understand everything together.* (TA11, BRR)

Code 14: Skipping (SKIP)

This coding category could also be termed as the lack of clarification. Readers omit thinking aloud about a word, phrase, clause, sentence, or section in the text. The reasons for skipping could be different – readers either do not understand something, or they may understand it but they want to move on.

> Text: Bankers hope to bundle this with other loans to entertainers and fully securitize it, but this is proving hard to do.
>
> Think-aloud: *Bankers would like to join* ***this*** *together, but* ***this*** *is difficult.* (TA3, BRR)

Code 15: Generalization (GEN)

Readers use relative or demonstrative pronouns and umbrella terms like 'thing' or 'matter' to avoid specific reference to a word or phrase in the text either because they do not know it or for some other reason.

> Text: Equity investments
> Think-aloud: ***Some*** *investments* (TA14, SS)

Code 16: Monitoring comment (MONCOM)

Readers express their perceptions of the reading process and comprehension while making processing decisions. They comment on the text and how they process it. They may comment on how well they comprehend the text or what difficulties they encounter and how they deal with them.

> Think-aloud: *I got lost.* (TA18, BRR)
>
> Think-aloud: *So, I'm slowly understanding what 'bond' means.* (TA8, SS)

Code 17: Using English within a Slovene think aloud (EN)

Readers think aloud in Slovene, but they intertwine their Slovene TAs with words, phrases, clauses, or sentences in English.

> Text: Moody's, a rating agency is considering downgrading Bowie bonds.
> Think-aloud (Slovene): *Ta Moody's, ta '****rating****' agencija premišljuje, da bo '****downgrading****', se prav, degradirala te Bowijeve obveznice.* (TA11, SS)
>
> Think-aloud (English): *This Moody's, this'****rating****' agency is thinking about '****downgrading****'; that is degrading Bowie's bonds.* (TA11, SS)

6.4 Analysis of the think-aloud data

The first step in the analysis of the think-aloud data involved a tabulation of coding frequencies and scores in tests of reading comprehension, grammar knowledge and background knowledge for each reader. The aim of the analysis of the think-aloud data and test results was to address the following research questions:

RQ 7: How do readers with high or low Background Knowledge read subject-specific texts?

Subquestion: What are the differences in patterns/strategies that readers with high or low BK use in the process of RC of subject-specific texts?

The answers to these questions were aimed to provide evidence to link qualitative (i.e. think-aloud) results to the results of the quantitative (i.e. testing) study.

Table 6.1 presents 17 think-aloud coding categories (i.e. 1.CT-17.EN) and their frequencies in the think-aloud protocols of the 24 students (i.e. TA1-TA24). Four sets of protocols marked with an asterisk (*) were incomplete or missing (i.e. TA6, TA19, TA20, TA24), as these students did not complete the think-aloud session and this affected their results. In addition, the lower part of Table 6.1 provides: a) means of reading comprehension, grammar knowledge and background knowledge tests for all readers, and b) individual scores of reading comprehension, grammar knowledge and background knowledge tests for 24 readers.

Table 6.1 Frequency of think-aloud coding categories per reader TA1-TA24

Reader	All TAs	TA 1	TA 2	TA 3	TA 4	TA 5	TA 6 *	TA 7	TA 8	TA 9	TA 10	TA 11	TA 12	TA 13	TA 14	TA 15	TA 16	TA 17	TA 18	TA 19 *	TA 20 *	TA 21	TA 22	TA 23	TA 24 *
1. CP	792	27	13	32	47	31	7	53	21	21	66	42	34	89	24	20	39	68	21	21	5	37	14	39	21
2. AP	746	10	25	30	29	31	21	42	19	40	40	41	30	36	50	53	33	23	34	18	7	45	41	28	20
3. WP	474	1	39	25	26	31	19	14	7	22	11	20	24	20	63	36	11	10	28	16	3	16	21	5	6
4. CT	124	1	21	9	6	4		3	9	13	8	11	7	4	8	1	2	3	1	1		2	1	6	3
5. AT	42		3	5	1	2		6	1	3	2	4	5	1	5				1	1		2			
6. WT	185	3	19	5	11	8	2	15	4	20	6	8	9	5	30	6	6	3	7	1		12		3	2
7. ELAB	169	9	4	3	3	7	6	1	8	11	13	3	17	28	11	2	19	2	2	3			1	15	1
8. EVAL	43			1				1	7		9		2	10	2		5		1					5	
9. GUES	32			1				2	1	2	4		2	1	4	10					1		1	2	1
10. INFR	330	18	30	22	13	13	5	13	22	10	29	10	27	27	23	12	26	3	4	1		3		16	3
11. REP	65		6	1	4	7		3	7	7	1		1	3	12	6		2				2		2	1
12. SUW	175	6	24	8	15	9		3	4	9	5	13	6	2	33	3		1	9	4		14	3	3	1
13. SLU	253	2	1	7	11	13	1	9	8	7	17	22	6	5	37	16	8	1	10	20	13	15	12	4	8
14. SKIP	340	28	27	15	20	16	11	14	15	16	13	14	12	8	6	12	19	8	10	21	1	10	29	6	9
15. GEN	147	2	10	5	2	2	2	4	3	5	7	7	11	3	23	10	3	5	16	4	1	8	6	3	5
16. MONC	199	8	3	13	11	4	3	4	42	2	28	7	1	7	24	7	4		3	3		6	3	10	6
17. EN	1466	4	76	134	58	34	3	130	20	189	64	239	35	21	136	41	12	7	16	28	12	99	27	40	41
RC % (mean 55.9)		**45**	**54**	**76**	**45**	**74**	**47**	**51**	**69**	**51**	**78**	**33**	**45**	**69**	**47**	**47**	**63**	**69**	**39**	**71**	**74**	**57**	**35**	**41**	**63**
GK % (mean 61.4)		**80**	**67**	**53**	**27**	**57**	**57**	**77**	**43**	**57**	**73**	**87**	**57**	**93**	**57**	**50**	**90**	**37**	**50**	**47**	**43**	**63**	**57**	**80**	**73**
BK % (mean 15.7)		**17**	**9**	**2**	**7**	**18**	**24**	**14**	**14**	**5**	**40**	**9**	**12**	**17**	**2**	**7**	**33**	**31**	**24**	**17**	**9**	**14**	**7**	**33**	**9**

Green – highest scores in group Blue – lowest scores in group ***** missing/incomplete protocol

6.5 Forming high and low background knowledge groups

To compare differences between students with different levels of background knowledge, two groups of readers were formed (Table 6.2). The high background knowledge group consisted of four students (i.e. TA10, TA16, TA17 and TA23), whose background knowledge test scores were the best out of 24 and they ranged from 31-40%, and are shaded in green. The low background knowledge group comprised six students (i.e. TA3, TA4, TA9, TA14, TA15 and TA22), who had lowest background knowledge test scores out of 24, ranging from 2-7%, and are shaded in blue. Due to the distribution of scores, it was not possible to make the high and low background knowledge groups with the same number of students. However, it was still possible to make two clearly separate groups of readers as regards distinctly different levels of background knowledge. There was a group of 14 readers between the high and low background knowledge groups whose background knowledge scores ranged from 9-24%. This group clearly sets apart the high and the low background knowledge groups.

Table 6.2 Division into the high and low background knowledge groups

	Students	BK test score %	**RC test score %**	GK test score %
High BK group	**TA10**	40	**78**	73
	TA16	33	**63**	90
	TA17	31	**69**	36
	TA23	33	**41**	80
	Mean HBK group (N 4)	**34.25**	**62.75**	**69.75**
Low BK group	**TA3**	2	**76**	53
	TA4	7	**45**	26
	TA9	5	**51**	56
	TA14	2	**47**	56
	TA15	7	**47**	50
	TA22	7	**35**	56
	Mean LBK group (N 6)	**5**	**50.16**	**49.5**
	Mean all TAs (N 24)	**15.7**	**55.9**	**61.4**

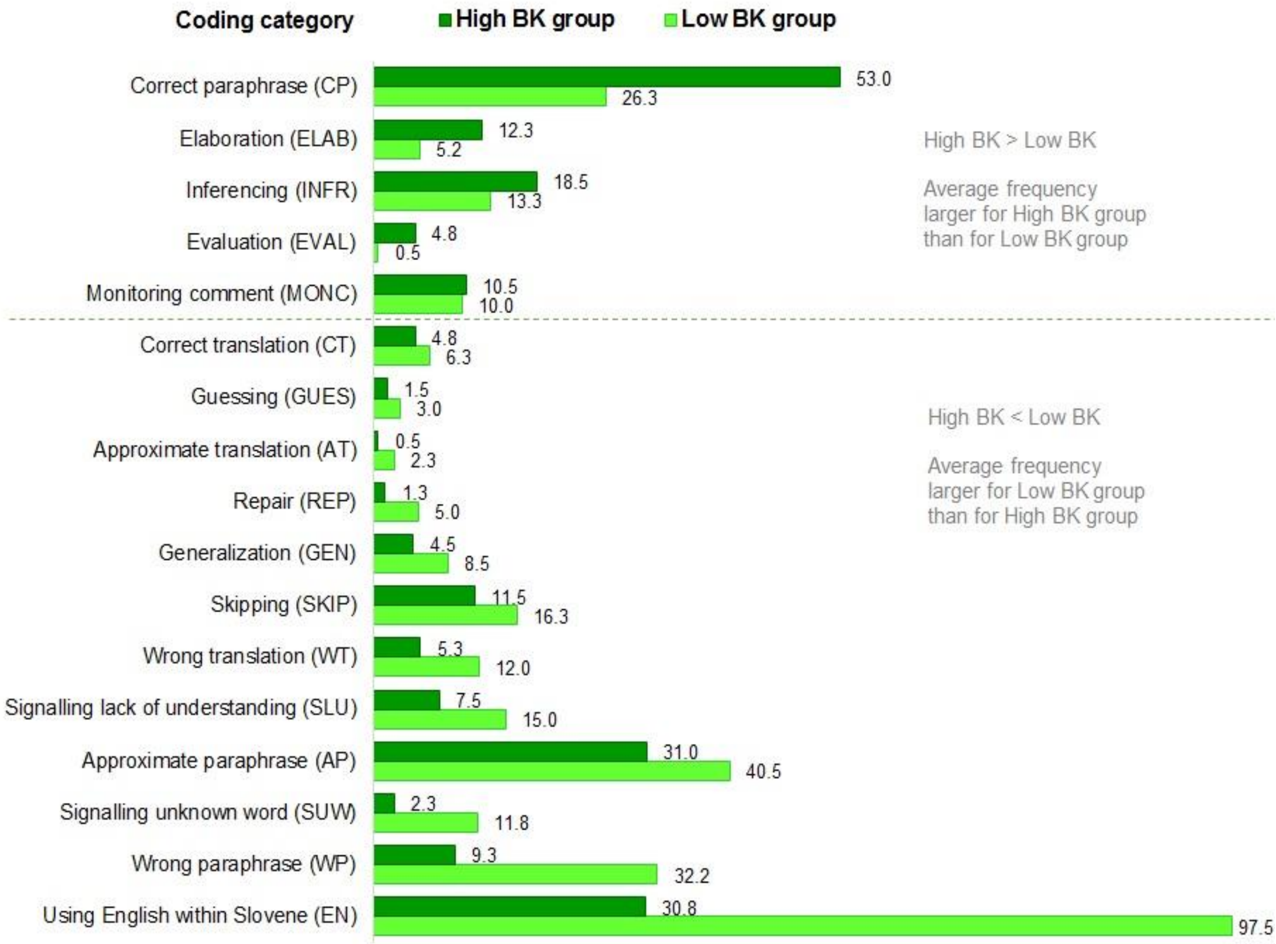
6.6 Reading comprehension scores of the high background knowledge and the low background knowledge groups

Despite the small number of students in the think-aloud study, it seemed worthwhile to see how reading comprehension scores relate to the use of coding categories. Apart from the division into two groups with different level of background knowledge, Table 6.2 shows how high and low background knowledge groups compare as regards their reading comprehension scores. It is clear that the high background knowledge group outperformed the low background knowledge group in reading comprehension. Three out of four students in the high background knowledge group achieved reading comprehension scores well above the mean score of 55.9%, ranging from 63-78%, and reader TA10 achieved the best reading comprehension score in the whole group (78%). There was one outlier in the group: reader TA23 had reading comprehension score below the mean (41%) and below the result of the majority of the low background knowledge group. In the low background knowledge group, five out of six students achieved reading comprehension scores below the mean score, ranging from 35-45%. As in the high background knowledge group, there was also one outlier: reader TA3, who had a reading comprehension score of 76%, which is much higher than other readers in the group. Despite both outliers in the high and the low background knowledge groups, the trend of reading comprehension scores is obvious. Readers with better background knowledge outperformed readers with lower background knowledge. This finding is parallel to the result in the testing study, which showed a positive correlation of reading comprehension scores with background knowledge scores. The better the background knowledge of students, the better their reading comprehension.

6.7 Differences in processing patterns between high background knowledge and low background knowledge groups

The next step after forming the high and low background knowledge groups was to calculate and compare the frequencies of think-aloud categories used by students in the two groups. Figure 6.1 shows the average coding frequencies for both groups in parallel bar charts. Figure 6.2 highlights the differences in coding frequencies between the two groups. This section summarizes the dominant trends in both groups and provides answers to RQs 7 and 8.

Figure 6.1: Average frequency of think-aloud coding categories in high BK group and low BK group

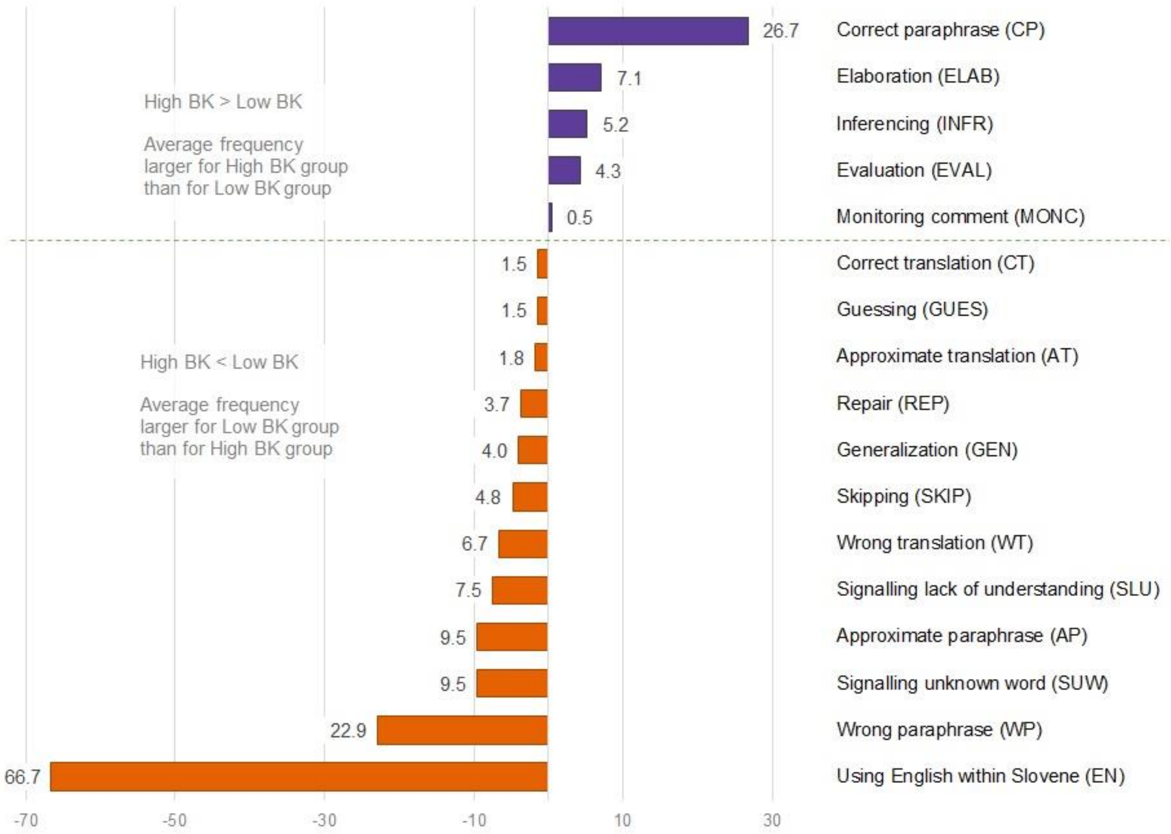


1. **More correct paraphrasing and better reading comprehension test scores in the high background knowledge group**

   The high background knowledge group's better reading comprehension is shown not only by their better reading comprehension test scores compared to the low background knowledge group (means: 62.75 vs 50.16) shown in Table 6.2, but also by the fact that correct paraphrasing (CP 53) ranks as the most frequently used coding category in the high background knowledge group (Figure 6.1). Correct paraphrasing (CP) was used 100% more by the high background knowledge group compared to the low background knowledge group (53 high BK vs 26.3 low BK), which clearly suggests better processing in the high background knowledge group. This evidence corroborates the findings of prior studies that background knowledge correlates with reading comprehension and that topic familiarity has a positive effect on comprehension (Chang, 2006; Bengeleil & Paribakht, 2004; Bråten & Strømsø, 2003; Pulido, 2004, Leeser, 2007; Cromley & Willis, 2016).

   Approximate paraphrasing (AP) was the third most frequently coded category in both groups, but it had a different absolute frequency: 40.5 for the low background knowledge and 31 for the high background knowledge group. This further suggests better comprehension in the high background knowledge group. The relationship between higher background knowledge and higher reading comprehension may not be causal, but the majority of reading comprehension test scores also show that readers with better subject knowledge outperformed students with less of it.

Figure 6.2: Differences in average frequencies of think-aloud coding categories between high and low background knowledge groups



2. **More elaborations, inferences and evaluations in the high background knowledge group**

   Figure 6.2 shows differences in average frequencies of the think-aloud coding categories between the two groups with different background knowledge. Five coding categories top right were more frequent in the high background knowledge group, whereas all the remaining twelve categories in the lower left part of graph were more frequent in the low background knowledge group. Apart from correct paraphrasing (CP), high background knowledge readers generated more inferences (INFR), elaborations (ELAB) and evaluations (EVAL) than low background knowledge readers. Using these strategies is evidence of readers' use of their background knowledge when creating the situation model of text comprehension. According to Grabe (2009), *inferencing* is a basic cognitive process in interpreting the environment. When reading, it involves "connecting the text with our memory resources that provide our background knowledge" (Grabe, 2009, p. 68). Similarly, other authors have defined inferencing as external evidence of the interaction between the reader and the text, which shows what knowledge the reader brings to the text (Hammadou, 2000; Barry & Lazarte, 1998; Nassaji, 2004). Grabe (2009) distinguishes bridging and elaborating inferences that may be used for several purposes including: a) integrating new information with background knowledge, b) interpreting decontextualized information, c) synthesizing information from multiple sources, d) evaluating information in terms of readers' goals and attitudes, and finally e) understanding conceptually different information. The success of inferencing was found to depend on two elements: background knowledge and textual clues (Hu & Nassaji, 2014; 2012; Cromley & Willis, 2016). Chang (2006) found that readers generated less inferencing if they were reading on an unfamiliar topic and concluded that inferencing is facilitated by topic familiarity. This finding is supported by the results in the present study as the high background knowledge readers produced not only more inferences but also more elaborations and evaluations than the low background knowledge readers. However, it is worth pointing out that there was great individual variation among students in terms of how much they inferred, elaborated, or evaluated. This may reflect the differences in readers' depth of background knowledge as well as the differences in readers' personality traits.

Inferences are related to *elaborations;* some authors actually list elaborations as a subcategory of inference and refer to them as "elaborating inferences" (Van Dijik & Kintsch, 1983; Grabe, 2009). A closer look at the elaborations used by readers in this study showed a dichotomy in the type of background knowledge readers drew on and the relevance of elaborations in a particular context. Some elaborations were clearly relevant to the text and showed that readers relied on the right background knowledge that improved comprehension. If readers retrieved contextually relevant information from their long-term memory, they used it to analyse the text in their working memory and this facilitated their comprehension. Conversely, some elaborations were completely unrelated to the text and they did not appear to affect comprehension. In either case, relevant or not, elaborations show the manner in which readers integrate text information into their existing background knowledge. But it is the congruence between the text base and readers' background knowledge that facilitates better comprehension.

3. **More local focus in the low background knowledge group (CT, AT, WT)**

   The low background knowledge readers focused approximately 100% more on individual words than the high background knowledge readers (20.6 low BK vs 10.6 high BK) as indicated by the total of correct, approximate, and wrong translations. Out of 20.6 word translations in total, the low background knowledge readers produced the majority of wrong translations (12) and these results along with their more pronounced focus on individual words suggest comprehension difficulties compared to the high background knowledge readers. It appears that these problems resulted primarily from a combination of two elements: readers' lack of vocabulary and lack of background knowledge.

4. **More signalling unknown words, lack of understanding, skipping, and generalizations in the low background knowledge group (SLU, SKIP, GEN)**

   The low background knowledge readers signalled lack of understanding of words or sections of the text much more than the high knowledge readers (26.5 low BK vs 9.8 high BK). The low background knowledge readers also skipped parts of the text more (16.5 low BK vs 11.5 high BK) and used more generalizations (8.5 low BK vs 4.5 high BK) than the high background knowledge readers. Signalling unknown words and lack of understanding shows readers' awareness of their problems in the comprehension process. It is worth noting that the low background knowledge readers kept more focus on the unfamiliar elements than high knowledge readers.

5. **More use of English within think-alouds in Slovene and more repair in the low background knowledge group (EN, REP)**

   Although both groups used a substantial amount of English within their think-alouds in Slovene, the low background knowledge readers used English more frequently, actually by far the most frequently of all coding categories, outnumbering the high background knowledge group by 200% (97.5 low BK vs 30.8 high BK). The think-aloud data suggest that readers switched from Slovene to English for a variety of reasons. They may have failed to understand a word or a phrase, they may have not found the word, phrase or sentence of particular importance for comprehension, or they may have wanted simply to speed up reading. The use of English within a verbal protocol in Slovene also appears to reflect the jargon of young Slovenes, which abounds in English words. The low background knowledge readers also tried to repair their understanding more often than the high background knowledge readers (5 low BK vs 1.3 high BK).

6. **Similar degree of total paraphrasing (CP + AP + WP) and monitoring (MONC) in both background knowledge groups, but differences within each category (CP, AP, WP)**

   The frequency of all paraphrasing coding categories including correct paraphrases (CP), approximate paraphrases (AP) and wrong paraphrases (WP) was similar in both groups (95.3 high BK vs 98.7 low BK), however the breakdown per category differs. There were more wrong paraphrases in the low background knowledge group (32.2 low BK vs 9.3 high BK) and more correct paraphrases in high background knowledge group (53 high BK vs 26.3 low BK) as referred to in point 1. Despite the similar total quantity of paraphrasing in both groups, more correct paraphrasing and less wrong paraphrasing in high background knowledge group suggests superior comprehension of the high background knowledge readers. Finally, the low and high background knowledge readers used approximately the same amount of monitoring comments (10.5 high BK vs. 10 low BK).

6.8 Chapter summary and conclusions

This chapter addressed two research questions. It explored how readers with different levels of background knowledge read subject-specific texts. In particular, it examined the strategies and patterns readers use in the process of comprehension and how they build their model of text interpretation. By using the think-aloud methodology, readers' processing patterns were elicited and analysed in relation to the text and deployment of readers' knowledge. The analysis of verbal protocols indicated how readers are challenged by reading subject-specific texts and how they fill lexical or conceptual gaps in texts in order to make a coherent mental model of text comprehension.

Readers' background knowledge was found to facilitate reading comprehension if it was congruent with the text base, and if it was contextually relevant. Readers were divided into the high and the low background knowledge group based on their background knowledge test scores. Then the think-aloud data of the two groups were compared. The high background knowledge group used more correct paraphrasing, inferencing, elaboration, and evaluation, whereas the low background knowledge group showed more of a local and literal focus. They signalled unknown words or lack of understanding more frequently than the high background knowledge readers. More often they skipped parts of text, used generalizations and English within in their think-aloud protocols in Slovene. Low background knowledge readers were also less flexible and less successful in building a coherent text model. Finally, a comparison of reading comprehension test scores of the three subject-specific texts showed that high background knowledge group outperformed the low background knowledge group.

CHAPTER 7
DISCUSSION

7.1 Chapter aim and overview

This chapter discusses the study results with regard to the aims, objectives, and research questions it set out to investigate. It considers the contribution of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity to L2 Reading Comprehension of subject-specific texts by combining the results of quantitative and qualitative studies and discusses them in light of evidence from previous theory and research studies.

The chapter opens with a discussion of the overall results regarding the effect of Background Knowledge on Reading Comprehension, in particular its effect in reader subgroups with different levels of either Grammar Knowledge or Background Knowledge. This section is followed by a discussion of the findings related to the difference between two operationalizations of background knowledge: tested Background Knowledge (BK) and Self-reported Familiarity (SRF). It goes on to discuss the relationship between Background Knowledge and Grammar Knowledge by examining the results of testing the threshold and compensation effects. The discussion of quantitative results is complemented by a discussion of qualitative results on how background knowledge is deployed in the L2 reading process. Finally, it discusses the results that concern readers' assessments and comments in the post-reading questionnaire, including readers' perceptions of four factors: text difficulty, topic familiarity, topic interest, and their reading habits. The chapter closes with an overview of the main conclusions related to the questions raised and the outcomes of the study.

7.2 The effect of Background Knowledge and Grammar Knowledge on Reading Comprehension (based on bivariate analysis)

This study investigated the role that readers' Background Knowledge plays in L2 reading of subject-specific texts and how it compares to Grammar Knowledge and readers' perception of their topic knowledge. This study's aim is reflected in RQs 1 and 2.

RQ 1: Do Background Knowledge, Grammar Knowledge, and Self-reported Familiarity affect Reading Comprehension of subject-specific texts?

RQ 2: To what extent do Background Knowledge, Grammar Knowledge and Self-reported Familiarity explain individual differences in Reading Comprehension of subject-specific texts?

This section presents a discussion of the results to answer RQs 1 and 2. The main aim was to determine the strength of relationship between three independent variables (Background Knowledge, Grammar Knowledge, Self-reported Familiarity) and one dependent variable (Reading Comprehension). In order to do that, data were obtained from 358 students who took part in the main testing study by reading three subject-specific texts (i.e. "Star Struck" (SS), "Bonds that Rock and Roll" (BRR), and "Taking Stock" (TS)). Before reading, students took tests of grammar knowledge and background knowledge; after reading each text, they answered comprehension questions and then filled in the perception questionnaire. Test scores and questionnaire data were subjected to a series of statistical analyses including bivariate correlation and multiple regression analyses. I will discuss the results of both analyses in two steps.

The key result rendered by bivariate correlation analyses showed a statistically significant relationship of moderate strength between Background Knowledge (BK) and Reading Comprehension (RC) ($rho_{BK} = .471^{**}$ at $p < .001$) (see Table 5.30). This relationship was slightly stronger than the relationship between Grammar Knowledge (GK) and Reading Comprehension ($rho_{GK} = .412^{**}$ at $p < .001$), which was also moderate and statistically significant. This finding suggests that both Background Knowledge and Grammar Knowledge are correlated with Reading Comprehension. This means that readers who possessed better background knowledge and grammar knowledge, were better at reading comprehension. The bivariate correlation analysis results provide partial answers to RQs 1 and 2 by showing the strength of the relationship between the variables; however, they do not allow for any predictions or causal conclusions to be made about the effect of the independent variables on the dependent one. Therefore, to obtain a complete answer to RQs 1 and 2 that could be compared to findings of other research studies, multiple regression analysis was performed. The results are discussed in Section 7.4, while the discussion of the Self-reported Familiarity variable compared to Background Knowledge follows next in Section 7.3.

7.3 Background Knowledge versus Self-reported Familiarity (based on bivariate analysis)

The third variable used in the formulation of RQs 1 and 2 was Self-reported Familiarity (SRF). This variable was introduced to allow a comparison between the tested background knowledge and perceived topic knowledge as reported by readers. The former was operationalized as a score on the test of background knowledge (BK) and the latter as readers' familiarity assessment on a 7-point scale (SRF). The intention was to determine whether the two operationalizations of background knowledge are predictors of the criterion variable Reading Comprehension and to test whether there is a distinction between them.

Readers' rating of their topic or content familiarity with the text has been one of the most frequently used measures of knowledge in L2 reading studies (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Lin, 2002; Brantmeier, 2003; Salmani-Nodoushan, 2003; Pulido, 2007; Eidswick, 2010; McNeil, 2011; Lahuerta Martinez, 2013; Shin et al., 2019). Despite their popularity in reading research, familiarity self-reports have received a fair amount of criticism. Song and Reynolds (2022) criticized self-reports for being based on assumptions rather than being a controlled measure, which is something especially needed when studying reading comprehension of expository texts. Lin (2002, p. 187) noted that readers' perceptions of their background knowledge may "deviate from their actual performance" and "fail to provide a true account" of their knowledge. Acknowledging the methodological problems involved in self-reporting led Uso-Juan (2006) to conclude: "The best way to gain a complete picture of students discipline-related knowledge is to give them knowledge tests in the subject areas of the intended research." In her study of L2 reading, Uso-Juan used background knowledge test to assess readers' discipline knowledge. Similarly, recent L2 reading studies have turned to testing background knowledge (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020). Despite this shift in the operationalization and measurement of background knowledge towards testing knowledge, the effect of either tested or perceived background knowledge on reading comprehension has not been compared in one single study. This motivated the inclusion of two different operationalizations of background knowledge in the present study (i.e. BK and SRF), which was intended to enable a comparison of their relationship with Reading Comprehension.

Bivariate correlation analysis was used in this study to compare Background Knowledge and Self-reported Familiarity with regard to their association with Reading Comprehension. The results revealed a more robust relationship between Background Knowledge and Reading Comprehension (rho_{BK} = .471** at $p < .001$) than the relationship between Self-reported Familiarity and Reading Comprehension (rho_{SRF} = .162** at $p = .002$), while correlations of both were statistically significant (see Table 5.30). This finding indicates that Background Knowledge was more strongly associated with Reading Comprehension than Self-reported Familiarity, which provides further information that informs the answers to RQs 1 and 2 (described in Section 7.2). However, to obtain complete answers about the differences in predictive values of both variables, multiple regression was used and its results are discussed next in Section 7.4.

To further explore the difference between Background knowledge and Self-reported Familiarity, the two variables were not only compared in terms of how each relates to Reading Comprehension, but also how they relate to each other. To test this relationship, bivariate correlation analysis was used. The result showed a weak but statistically significant relationship between Background Knowledge and Self-reported Familiarity (*rho* = .196** at $p < .001$) (see Table 5.30). This finding suggests that both Background Knowledge and Self-reported Knowledge appear to pull in the same direction, but their association is small. This finding may point to several conclusions. First, self-reporting may not be a substitute for background knowledge testing, particularly when discipline-related knowledge is concerned. Although they seem to share a portion of the same construct, tested Background Knowledge and Self-reported Familiarity cannot be taken as interchangeable measures. These study results suggest that familiarity self-reports cannot be taken on a par with discipline-related background knowledge tests, because the weak association between the tested Background Knowledge and Self-reported Familiarity may contribute to doubt in the validity of self-reporting one's knowledge as a measure of background knowledge. This concern was expressed by Cervetti and Wright (2020), who also point to the critical fact that reading studies "have largely relied on familiarity ratings or experiential indicators of knowledge, rather than direct assessments of knowledge" (Cervetti & Wright, 2020, p. 248).

This study's finding of a stronger association between Background knowledge and Reading Comprehension in comparison to Self-reported Familiarity and Reading Comprehension may cast additional doubt on familiarity reporting as a measure of background knowledge. It also corroborates Uso-Juan's (2006) claim that a tested assessment of background knowledge may be a better indicator of readers' knowledge than self-reporting. Consistent with her observations, this study offers empirical evidence to support the position that there is a significant difference between tested Background Knowledge and Self-reported Familiarity in relation to Reading Comprehension. This may be because self-reports are based on individuals' perceptions of their own knowledge, which are prone to subjectivity.

It is worth noting that I explore the distinction between background knowledge and familiarity not only by how they are measured, either through testing or reporting, but also by what is being measured. It appears that knowledge and familiarity differ in terms of breadth and depth. Chou (2011) stressed that the conceptualization of background knowledge is deeper than familiarity because knowledge not only contains surface information, but it also involves concepts and relationships between them. Therefore, background knowledge goes beyond familiarity, and it also involves the knowledge of terminology. Similarly, Guthrie (2004) contends that "background knowledge should represent interconnected concepts and information related to the topic." Such understanding of background knowledge can be linked to Nation's (2017) claim that an indicator of discipline knowledge is the technical vocabulary we learn as we learn the discipline. Background knowledge is therefore articulated through vocabulary, so vocabulary is a proxy for conceptual knowledge (Fisher & Fry, 2009). What transpires from the literature is that operationalization and measurement of background knowledge should be based on testing, and it should encompass the information, concepts, and relationships deemed central in the discipline. This resonates with the empirical evidence in this study, which supports the distinction between background knowledge and self-reporting with regard to their role in reading comprehension.

As the results discussed in Section 7.3 refer only to bivariate analysis involving the two variables, we need now to turn to the results of multivariate regression for a complete picture of effects of all predictor variables on Reading Comprehension.

7.4 An overall picture of the contribution of Background Knowledge, Self-reported Knowledge, and Grammar Knowledge to Reading Comprehension (based on multiple regression analysis)

Although bivariate correlation analysis can provide a picture of the strength of relationships among the studied variables, it cannot reveal their predictive value. This means that results of bivariate correlation presented in Section 7.2 do not provide full answers to RQs 1 and 2. In order to ascertain the predictors among the independent variables and answer RQs 1 and 2, multiple linear regression analysis was performed. The results of multiple linear regression analysis with Grammar Knowledge, Background Knowledge, and Self-reported Familiarity as predictors showed that 29.5% variance of the criterion variable Reading Comprehension can be explained by the contribution of the three independent variables in the regression model (see Table 5.31). The two independent variables Background Knowledge and Grammar Knowledge affected Reading Comprehension at a medium level, and the influence of both was statistically significant. More specifically, there was a slightly larger contribution of Background Knowledge (standardized $\beta = .343$ at $p < .001$) than Grammar Knowledge (standardized $\beta = .312$ at $p < .001$). The third independent variable Self-reported Familiarity (standardized $\beta = .071$ not significant at $p < .05$) did not have a statistically significant influence on Reading Comprehension. This finding leads to the important conclusion that Self-reported Familiarity cannot be considered a predictor of Reading Comprehension in this study, whereas Background Knowledge and Grammar Knowledge are both significant medium-strength predictors of Reading Comprehension. The results of multiple regression and bivariate analysis have now provided complete answers to RQs 1 and 2, and they offer empirical evidence for a better understanding of the relative contribution of three predictor variables to Reading Comprehension in this study.

To date, a number of studies investigating the role of background knowledge in L2 reading comprehension have found that both L2 proficiency and background knowledge have a facilitative role in comprehension. L2 proficiency was largely found to play a stronger role by explaining a larger portion of variance (Tan, 1990; Chen & Donin, 1998; Clapham, 1996; Uso-Juan, 2006; Karimi, 2017). In a report on the interaction between background knowledge and language proficiency affecting TOEFL reading performance, Hill and Liu (2012) highlight the contribution of both factors, but emphasize that studies with large sample sizes tended to find a stronger effect of L2 proficiency compared to discipline-related background knowledge.

These study results may offer some parallels to Clapham (1996) and Uso-Juan (2006), two comprehensive studies that have used larger sample sizes and analysed data sets by performing multiple regression. They have both found L2 proficiency and background knowledge to be predictors of reading comprehension, although with varying predictive values. Clapham (1996) reported that 26% of variance in reading comprehension was accounted by language proficiency and 12% by background knowledge, whereas Uso-Juan (2006) found that language proficiency accounted for 58–65 % of variance, while discipline-related background knowledge accounted for 21–31%. The results in this study are similar to Clapham with regard to total variance explained, the beta coefficients show that Background Knowledge tends to be a significant moderate predictor of Reading Comprehension that is slightly stronger than Grammar Knowledge.

Both Clapham (1996) and Uso-Juan (2006) found interaction between background knowledge and L2 proficiency. Clapham (1996) reported that background knowledge had an effect for intermediate proficiency students, in her study set at proficiency scores from 60–80%. Readers outside this range, either higher or lower, did not appear to benefit from their background knowledge, which she interpreted as a confirmation of lower and upper language thresholds for background knowledge effect. Similarly, Uso-Juan (2006) and Karimi (2017) concluded that language proficiency and background knowledge both play a key role in comprehension, but this role is particularly important in L2 reading and in multiple-text reading, in which comprehension is assessed across texts. Karimi (2017), concluded that the combination of discipline-specific background knowledge and L2 proficiency is more pronounced in L2 reading and in multiple-text comprehension rather than single-text comprehension as shown by the difference in explained variance (e.g. 39% vs 26%).

In line with findings of this study, a stronger facilitative role of background knowledge in L2 reading comprehension has been reported in several studies, especially the ones that explored discipline-related background knowledge (Lin & Chern, 2014; Hwang, 2019; Hwang & Duke, 2020). Lin and Chern (2014) found discipline-related knowledge to be a stronger predictor of reading comprehension measured by summary writing than L2 proficiency. In a study comparing L1 and L2 comprehension, Hwang (2019) found that science-domain knowledge was the strongest predictor of reading comprehension for both L1 and L2 reading. Similarly, Rydland et al. (2012) found prior knowledge to be the strongest predictor of reading comprehension in comparison to word decoding, L1 and L2 vocabulary, with vocabulary moderating the contribution of prior knowledge. Shin et al. (2019) determined that working memory capacity (WMC) moderates the effect of background knowledge. If readers had a better working memory capacity, this was beneficial for their ability to leverage background knowledge, which led to a more strategic approach and readers' ability to supress irrelevant information. Hwang and Duke (2020) found a strong association between science domain knowledge and reading comprehension that was stronger for L2 readers than for L1 readers.

Evidence contrary to this study's findings was provided in McNeil's (2011) study, in which the role of background knowledge was compared to reading strategies with regard to their contribution to L2 reading comprehension. Reading strategies turned out to be a considerably stronger predictor that accounted for 53% of variance in L2 reading, whereas background knowledge only accounted for .4%. As this was an exploratory study, the sample size was small (e.g. 11 readers), and the author reported that only literal questions were used in reading comprehension assessment, text topic was general, but the text was linguistically demanding so that it required strategy use. In a later study, McNeil (2012) also posited a compensatory model of L2 reading for readers of different proficiency levels, without specifying the relative contribution of background knowledge. For lower- and higher-proficiency readers, the model proposed shifting contributions of L2 language knowledge, L1 reading ability, strategic knowledge, and background knowledge.

7.5 Individual texts: the contributions of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity (based on multilevel regression analysis)

Related to RQs 1 and 2, a subquestion emerged in relation to individual texts used in the study (i.e. “Star Struck” (SS), “Bonds that Rock and Roll” (BRR), and “Taking Stock” (TS)). Specifically, it enquired whether the contribution of independent variables to Reading Comprehension varies with regard to individual texts.

Sub-question: Does the contribution of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity to Reading Comprehension of subject-specific texts vary among the three texts?

To address this question, a multilevel regression model was constructed (see Table 5.32). It showed a statistically significant moderate effect of both Background Knowledge and Grammar Knowledge on Reading Comprehension for each of the three texts. However, there were some differences among texts, as Grammar Knowledge and Background Knowledge had a stronger effect on Reading Comprehension for the BRR text than on Reading Comprehension for the SS or TS texts. There was also a small positive effect of Self-reported Familiarity for individual texts on Reading Comprehension of individual texts, which could not be observed for all texts combined in a multivariate setting. This evidence further confirms that the impact of Background Knowledge and Grammar Knowledge on Reading Comprehension did not just refer to all texts together, but it was also observed for individual texts. On the other hand, the same cannot be claimed for Self-reported Knowledge, as this variable had a small positive effect on Reading Comprehension of individual texts, but no such effect could be observed for all texts combined in a multivariate setting. Therefore, the influence of Self-reported Familiarity was found to be limited to individual texts (i.e. BRR only) and this could be understood as additional difference between Background Knowledge and Self-reported Knowledge, the two operationalizations of background knowledge used in this study. It may provide an extra argument for the claim that readers' assessment of topic familiarity with a particular text could not be compared with tested background knowledge with regard to its predictive power in Reading Comprehension.

7.6 The threshold effect (based on multiple level regression)

RQs 3 and 4 ask whether the impact of Background Knowledge on Reading Comprehension varies according to the level of readers' background knowledge and grammar knowledge.

RQ 3: Does the contribution of Background Knowledge to Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

RQ 4: To what extent does the contribution of Background Knowledge to Reading Comprehension of subject-specific texts vary among the subgroups of students divided according to their higher or lower Background Knowledge and Grammar Knowledge?

Hypothesis: Students can use their Background Knowledge only at a certain level of Grammar Knowledge

To answer the questions and test the hypothesis, a multiple regression model was constructed with readers divided in nine subgroups with regard to the level of their Grammar Knowledge and Background Knowledge (i.e. low, medium, high; 3x3 design; see Tables 5.33, 5.34 and Figure 5.7). Having explained 30.5% of the variance in Reading Comprehension scores, the model showed a statistically significant influence of Background Knowledge and Grammar Knowledge on Reading Comprehension, but no influence of Self-reported Familiarity (see Table 5.35), which is similar to the result in the initial regression model. The findings revealed that the effect of Background Knowledge on Reading Comprehension could be observed at all levels of Grammar Knowledge. There was no lower Grammar Knowledge limit below which Background knowledge effect would not be detected. Therefore, these findings do not seem to support the lower-level threshold hypothesis. However, the model revealed a slightly less prominent effect of Background Knowledge at the highest levels of Grammar Knowledge. Nevertheless, this finding could be not interpreted as a confirmation of the upper threshold. A possible explanation for these results in comparison to other studies that confirmed the lower threshold effect may lie in the distribution of grammar test scores which was skewed towards higher results. This might imply there were no "true" low-level grammar knowledge subjects or in other words, the readers in this study have already reached a L2 level that allows them to use background knowledge to benefit their comprehension. On the other hand, a relatively high level of grammar knowledge may also suggest that this result is difficult to compare with those studies that involved subjects with lower levels of grammar knowledge and detected the lower threshold effect.

The existence of lower and upper thresholds has been tested in several studies that wanted to determine either the level of L2 proficiency below which readers cannot use background knowledge effectively, or the point above which readers no longer rely on background knowledge. L2 reading studies have provided inconsistent empirical evidence for thresholds (Cervetti & Wright, 2020). Some studies detected the lower threshold (Clapham, 1996; Uso-Juan, 2005; Ridgway, 1997; Karimi, 2016; Shaw & McMillion, 2018), others found evidence for the upper threshold (Clapham, 1996), while the evidence from some studies was inconclusive (Krekeler, 2006).

Clapham (1996) confirmed both the lower and upper L2 proficiency thresholds by determining the cut off-level between 60–80% grammar test score. This means that readers below this level could not benefit from their background knowledge in their reading comprehension. On the other hand, readers above this level were so proficient that they did not need to draw on their background knowledge. This means that the background knowledge effect was manifest only between the two thresholds. Similarly, Uso-Juan (2006) tested thresholds, but detected only the lower threshold, acknowledging: "Despite having maximum discipline-related knowledge, learners need a linguistic threshold to be able to read an academic passage" (Uso-Juan, 2006, p. 220). Uso-Juan's finding was in accord with Ridgway (1997), who also found support for the lower threshold, but not for the upper threshold. He reported that results were mixed for different texts (i.e. business and built environment) due to the difficulty and opacity of texts. Apart from the varying degrees of text specificity, he also attributed the findings to a small sample size. Ridgway suggested that “While background knowledge is always at work, the effect is not always detectable. It appears that a reader must be at a certain level with relation to the text for the effect to be observable” and concluded that “There seems to be an intermediate area where background knowledge is a significant factor” (Ridgway, 1997, p. 161).

The results from this study do not offer support for thresholds and thus in this respect tend to go counter Clapham (1996), Uso-Juan (2006), and Ridgway (1997) as the effect of background knowledge on reading comprehension was observed at all proficiency levels. There was no evidence of cut-off points in proficiency levels that could be seen as thresholds for the background knowledge effect. In that respect, the empirical evidence from this study is more consistent with the findings of Krekeler (2006), who similarly did not find support for thresholds. In a study involving a mixed group of international students with German as L2, Krekeler (2006) tested background knowledge effect in LSAP tests, aiming to determine whether readers would be able to draw on their knowledge only at certain levels of proficiency. He found a strong effect of background knowledge on reading comprehension scores with some variation of the effect at different proficiency levels, but the effect was not so substantial that any threshold or cut-off point could be confirmed. Readers could use their background knowledge regardless of their level of L2 proficiency. He concluded that if thresholds exist, they are not sharp cut-off points but gradual and they depend on text specificity. Krekeler (2006) attributed the results of his study to the low variation in L2 proficiency and background knowledge levels among his readers as well as the relatively low subject specificity of texts. A similar observation regarding knowledge levels could be made for the readers of the present study. Readers’ grammar scores were more skewed to the higher levels, and background knowledge scores to the lower ones. This distribution has affected the regression results to some extent. On the other hand, the specificity of texts was controlled for in this study by involving subject experts in the selection of subject-specific texts, as suggested by Douglas (2000). In line with Krekeler (2006), the evidence in this study leads to the conclusion that Background Knowledge (operationalized as a BK test score) affects Reading Comprehension of subject-specific texts, at all levels of L2 Grammar Knowledge. Conversely, the same effect could not be claimed for Self-reported Familiarity.

7.7 The compensation effect (based on Kruskal-Wallis H test)

The possibility of a compensation effect was addressed in RQs 5 and 6 that asked whether students with better Background Knowledge (or better Grammar Knowledge) can compensate for lower Grammar Knowledge (or lower Background Knowledge).

RQ 5: Can students with better Background Knowledge compensate for lower Grammar Knowledge at Reading Comprehension of subject-specific texts?

RQ 6: Can students with better Grammar Knowledge compensate for lower Background Knowledge at Reading Comprehension of subject-specific texts?

Hypothesis: Better Grammar Knowledge allows students to compensate for lower Background Knowledge, and vice versa.

To answer the questions and test the hypothesis, the Kruskal-Wallis H test was used (see Table 5.36). Readers were first divided into nine groups with combinations of low, medium, and high levels of both Grammar Knowledge and Background Knowledge (i.e 3x3 design; see Figure 5.8). Of particular interest were the extreme groups, termed "compensating groups" (i.e. Low GK-High BK and High BK-Low BK) and the "extreme" group (i.e. Low GK-Low BK). When the Reading Comprehension score mean for Low BK-Low GK was compared to the Reading Comprehension mean of groups that either had high Background Knowledge or high Grammar Knowledge (i.e. High BK-Low GK, Low BK-High GK), in both cases the Reading Comprehension mean scores increased (from 28.48 to 35.20 or 34.52, respectively). This finding clearly showed that despite their lower Grammar Knowledge, readers who had higher Background Knowledge achieved better Reading Comprehension score than readers that had lower levels of both Background Knowledge and Grammar Knowledge (i.e. Low BK-Low GK group). This applied not only to the increase in the level of Background Knowledge, but also Grammar Knowledge. As illustrated by the gain in Reading Comprehension mean scores, the rise was slightly higher for Background Knowledge than Grammar Knowledge. As this result points to the compensation effect, it was then statistically tested by the Kruskal-Wallis H test. The outcome showed a statistically significant difference in Reading Comprehension scores of the compared groups at $p < 0.001$ (see Table 5.36). The "compensating groups" (i.e. Low GK-High BK group and High GK-Low BK group) reached higher Reading Comprehension scores than readers from other groups, with the exception of High GK-High BK group, thus providing evidence to confirm the compensation effect in this study.

The results of Kruskal-Wallis H test provide evidence to support the compensation hypothesis and offer answers to RQs 5 and 6. The compensatory effect was statistically significant, suggesting that readers with better background knowledge could make up for a lower level of grammar knowledge and vice versa. Generally, this means that when readers lack a particular source of knowledge, either language or content knowledge, they can leverage other resources available and make up for their deficiencies. The compensation effect found in this study can be linked to a number of other studies that have tested the compensation effect and found support for it (Koh, 1985; Chen & Donin, 1997; Al-Shumaimeri, 2006; Uso-Juan, 2006; Hwang & Duke, 2020). The findings of this study are consistent with Uso-Juan (2006), who obtained similar results by finding that compensation effect works in both ways for both language and content knowledge. She detected compensatory effect for those readers who lacked discipline-related background knowledge and had at least intermediate language proficiency. However, those readers whose L2 proficiency was lower could not do that and had to reach a language threshold in order to read academic texts effectively. Conversely, if readers had higher discipline-related knowledge, this could make up for their lower proficiency level. This result suggests that the compensation effect worked both ways for language and content knowledge, provided that at least an intermediate level was reached. Results corroborating a compensation effect were also reported by Al-Shumaimeri (2006) and Hwang and Duke (2020). Al-Shumaimeri observed that high proficiency readers had similar reading comprehension scores regardless of their background knowledge level and attributed them to the compensation effect. Hwang and Duke (2020) compared reading comprehension of L2 and monolingual readers and found that the compensation effect of science domain knowledge was particularly strong for L2 readers.

The findings in this study go counter to Lin and Chern (2014), who did not detect the compensation effect between L2 knowledge and background knowledge in reading comprehension of Taiwanese EFL learners. The ANOVA result showed that both background knowledge and language proficiency influenced reading comprehension, whereas background knowledge was a stronger predictor. The authors explained that absence of compensation may be due to the lexical and syntactic complexity of texts, readers' insufficient levels of background knowledge and language knowledge and their poor summary skills.

When discussing the results of this study in comparison to past studies of L2 reading, it is noteworthy to highlight variations in the study design that may affect the results. Three groups of factors tend to stand out: first, the operationalization and measurement of variables; second, subject specificity of texts; third, statistical analyses used. These factors may contribute to the inconsistencies in findings about knowledge-comprehension relationship (Cervetti & Wright, 2020; Shin, 2010). I have referred to operationalization and measurement of variables in sections 7.3 and 2.5. To recap, background knowledge in reading studies tends to be an umbrella term that may vary in depth and breadth, comprising general knowledge, discipline-related knowledge, cultural knowledge, and topic familiarity. Typically, in a large number of studies, the knowledge factor was not measured, but was either assumed by one's study discipline (Alderson & Urquhart, 1983; Koh, 1985; Peretz & Shoham, 1990; Ja'far, 1992; Clapham, 1996; Horiba & Fukaya, 2015) or self-reported topic familiarity (Afflerbach, 1990; Jensen & Hansen, 1995; Clapham, 1996; Khalifa, 1997; Lin, 2002; Brantmeier, 2003; Salmani-Nodoushan, 2003; Pulido, 2007; Eidswick, 2010; McNeil, 2011; Lahuerta Martinez, 2013; Shin et al., 2019) or more recently it has been assessed by using tests of knowledge (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020; Song & Reynolds, 2022).

Among textual factors, text-specificity has been noted to have a profound influence on whether the effect of background knowledge will be detected. Urquhart and Weir (1998) stated: "the more specific a text, the more important the contribution of background knowledge to comprehension, the less specific a text, the more important the contribution of language proficiency" (p. 144). However, despite acknowledging the importance of subject-specificity of reading texts, there has been little agreement on what makes a text highly specific (Clapham, 1996; Uso-Juan, 2006). In a discussion of factors that affect content specificity Douglas (2001) identifies the amount of field-specific vocabulary, the degree to which the specific purpose vocabulary is explained, the rhetorical functions of text sections, and the required knowledge of subject specific concepts. Apart from that, Gunderson et al. (2020) list more non-concrete items of vocabulary, more specialized and therefore difficult vocabulary, more vocabulary with non-standard meanings and generally more abstract material. They also note that content-specific materials contain more complex sentence structures, different kinds of visual aids, and more new material that is information-packed and concentrated. Nevertheless, while acknowledging the importance of specificity in LSP testing, it was highlighted that this concept integrates the language and content, general and specific language use, hence it is not possible to draw clear lines and separate language of different domains and general domain (O'Sullivan, 2012; Brunfaut, 2014).

Finally, the third group of factors that may have contributed to mixed results in L2 reading studies involving background knowledge are different statistical analyses used. Generally, they have ranged from correlation analyses to one- or two-way ANOVA and multiple regression. Uso-Juan (2006, p. 211) pointed out that studies that used these techniques "carried out the analyses by groups and not by individuals and thus missed a lot of information."

7.8 Qualitative aspects of the Background Knowledge effect on Reading Comprehension (based on think-aloud data)

Qualitative aspects of the effect of Background Knowledge on Reading Comprehension were addressed in RQ 7. Specifically, it looked into how readers with high or low Background Knowledge read subject-specific texts and what processing patterns and strategies they use in the process of reading comprehension.

RQ 7: How do readers with high or low Background Knowledge read subject-specific texts?

Subquestion: What are the differences in patterns/strategies that readers with high or low Background Knowledge use in the process of Reading Comprehension of subject-specific texts?

To answer RQ 7, a think-aloud study was conducted with 24 students (described in Chapters 4 and 6). Before thinking aloud while reading three texts, readers took the same grammar and background knowledge tests as the readers in the testing study. After reading each text, they filled in the perception questionnaire asking them to rate their topic familiarity, text difficulty, and topic interest. Verbal protocols were transcribed and coded, using a data-driven coding scheme. Having counted and validated the frequencies of coding categories in protocols (see Table 6.1), it was possible to determine the processing patterns and strategies used by readers while thinking aloud. The objective was to compare the results between readers with high and low Background Knowledge and determine any differences between them. After analysing the think-aloud data, it was possible to draw a distinction between the two groups by interest (i.e. High BK group and Low BK group), which consisted of four and six students respectively (see Table 6.2). The High BK group had a better mean Reading Comprehension score than the Low BK group (i.e. mean 63% vs 50%). This finding obtained from the think-aloud data is consistent with the results from the testing study and additionally confirms a positive impact of Background Knowledge on Reading Comprehension. Comparing the frequencies of think-aloud coding categories showed that the High BK group used more correct paraphrases, elaborations, inferences, and evaluations (see Figure 6.1), a finding similar to Mikeska (2010). In contrast, the Low BK group manifested more of a local focus by more frequent use of translations of individual words that were either correct, approximate, or wrong. This result echoes Kroner's (2014) observation about readers in his study who focused on understanding specific single words or sentences. Their attention to local coherence diminished the global coherence of their text representation and led to the failure to identify the important information in the text. In the present study, this local focus may have been related to readers' lack of vocabulary or background knowledge. This is consistent with Chou (2013) and Nassaji (2004), who found a relationship between vocabulary depth and strategy use, particularly inferencing. In contrast to the High BK group in this study, the Low BK readers more often signalled unknown words or lack of understanding. They used more generalizations and repair, and they more frequently resorted to using English during thinking aloud in L1. The emerging patterns and strategies may be a reflection of the processing difficulties they encountered while reading. It appeared that Low BK readers were aware of their comprehension problems, but they remained more focused on unfamiliar and difficult items compared to High BK readers. The coding categories that were used with similar frequency by the High and Low BK groups were the monitoring comments and total paraphrasing. Overall, the differences between high and low knowledge readers suggest that background knowledge was related to strategy use although there were great individual variations, a result that is in accord with Comer (2012).

7.9 Readers' perception of text Difficulty, Familiarity, and Interest and reading habits (based on post-reading questionnaire ratings and comments)

As one of the aims of this study was to compare the contribution of Background Knowledge and Self-reported Familiarity to Reading Comprehension, a perception questionnaire was used to obtain readers' assessment of their topic familiarity related to the three texts (i.e. SRF). In addition to the variables referred to in the research questions, the post-reading questionnaire aimed to obtain readers' perceptions about two additional factors: text difficulty and topic interest. Readers were asked to provide their rating on a 7-point scale and then answer open questions about their attributions of difficulty and familiarity. In addition, readers were asked to give some baseline data regarding gender and reading habits. The data obtained from post-reading questionnaires were subject to bivariate analyses together with readers reading comprehension scores for all texts together as well as for individual texts. Readers' answers to open questions were tallied and analysed, so that trends could be determined. The discussion of results is presented in Sections 7.9.1–7.9.4.

7.9.1 Aggregate results for all texts: Self-reported Familiarity, Difficulty, Interest (after bivariate analysis of post-reading questionnaire data)

To explore if there is a relationship between Reading Comprehension, Familiarity, Difficulty, and Interest, bivariate analysis was conducted for individual texts as well as all texts together (see Tables 5.37, 5.38, and 5.39). The results of bivariate analysis for all three texts together showed statistically significant correlations of all variables (i.e. SRF, Difficulty, Interest) with Reading Comprehension. A moderate negative relationship was found between Difficulty and RC (*rho* = −.308**, $p < .001$), which shows that readers who found texts more difficult had lower Reading Comprehension scores (see Table 5.39). This suggests that their assessments of difficulty were indicative of their test scores and thus correct. Conversely, Peretz and Shoham (1990) found that readers' perception of text difficulty was not a reliable predictor of their comprehension.

A weak statistically significant relationship was found between Reading Comprehension and Self-reported Familiarity (*rho* = .162**, $p = .002$) as well as between Reading Comprehension and Interest (*rho* = .192**, $p = .001$). Both results imply that readers who reported to be more familiar with the text topics or found text topics interesting performed slightly better on the Reading Comprehension test, though both correlations are weak. The first one, between Self-reported Familiarity and Reading Comprehension actually corroborates the findings from the testing study and was therefore not surprising. It confirms that there is a distinction between Self-reported Familiarity and Background Knowledge with regard to the strength of their relationship with Reading Comprehension.

The third analysed relationship was the one between Interest and Reading Comprehension. The correlation is low and indicates a weak relationship, albeit slightly stronger than between Self-reported Familiarity and Reading Comprehension. The role of interest has been investigated in several studies and a number of them have reported that interest is positively associated with reading outcomes or it has a positive effect on reading (Lahuerta Martinez, 2013; Kelly, 2014; Eidswick, 2010; Fox, 2020). In her discussion of affective individual differences, Fox (2020) compared the findings of studies that investigated content-related, topic-related and situational interest. The situational interest, described as experienced interest in the text topic after reading, was found as the best predictor of reading outcomes. In contrast, the findings in this study showed that situational interest did not have such a strong connection with Reading Comprehension.

Some other statistically significant relationships among variables have emerged from the bivariate analysis. The one that stands out is a moderate negative correlation between Self-Reported Familiarity and Difficulty (*rho* = −.347**, $p < .001$), which suggests an inversely proportional relationship between the two variables. Readers who were more familiar with the text topics found the texts less difficult. In addition, a moderate negative association was found between Interest and Difficulty (*rho* = −.423**, $p < .001$), indicating that readers who found texts more interesting were less likely to find them difficult. Finally, there was also a moderate positive association between Interest and Familiarity (*rho* = .405**, $p < .001$), suggesting that readers who found texts more interesting also found them more familiar. This outcome is contrary to that of Erçetin (2010) and Carrell and Wise (1998) who did not find any significant relationship between topic interest and prior knowledge.

7.9.2 Familiarity, Difficulty, and Interest ratings for individual texts

Readers' ratings of Familiarity, Difficulty, and Interest were tallied to determine the differences among the three texts (i.e. SS, BRR, TS). All text topics were found to be relatively unfamiliar by readers: BRR by 81% readers, SS by 73%, and TS by 57% (see Tables 5.13, 5.14, 5.15, and 5.16). However, about half of the readers rated that this lack of familiarity had a low effect on their reading comprehension, whereas around 20% reported that it had a high effect and 20% a medium effect. SS turned out to be the most interesting (by 44%) and the easiest of texts (by 50%). On the other hand, BRR was found to be the most difficult (by 57%) and most unfamiliar (by 81%) text. This assessment of BRR was supported by students' comments in open questions about texts. TS was found to be the most familiar of the three texts (by 24%), among the relatively low familiarity ratings for all texts. Readers perceived the effect of familiarity to be strongest for TS (by 37%), followed by BRR (by 31%) and SS (by 29%). Still, 50–41% reported low effect of familiarity.

7.9.3 Readers' comments about Familiarity, Difficulty, and Interest

To obtain a clearer picture about readers' perception of their Familiarity, sources of Difficulty, and Interest, readers were asked to provide comments on their ratings in the post-reading questionnaire. The tallied results showed some interesting trends for all texts (see Table 5.19). Overall, readers reported the main source of Difficulty to be vocabulary, described either as unknown, known, specialist, non-specialist, or difficult. This was followed by text (difficult, easy, long, or short) and topic (familiar, unfamiliar, interesting, or uninteresting). When asked about how familiarity or lack of it affected their reading, readers mainly referred to the reading speed (higher or lower), the level of understanding (easier, more difficult, easier to make connections, infer, predict), vocabulary knowledge (or lack of it) and finally, the text topic (interesting, uninteresting, familiar, or unfamiliar). Vocabulary, which turns out to be the key source of difficulty in readers' comments, is discussed in Section 7.11.

In reference to individual texts (see Table 5.21), TS was found to have the most difficult vocabulary by 50% of readers. BRR followed with difficult vocabulary found by 36% and text length by 21%. In contrast, SS was found to have difficult vocabulary by 28%, known vocabulary by 18% and it was also found to be an easy text by 18%.

7.9.4 Reading habits

In post-reading questionnaires, students also reported on their reading habits, both relating to their general reading and course-related reading. The data were subjected to bivariate analysis and generated the following findings (see Table 5.40). "General Reading in English" correlates moderately with Grammar Knowledge, Reading Comprehension, and Background Knowledge (rho_{GK} = .386**, rho_{RC} = .355**, rho_{BK} = .266**, all at $p < .001$), whereas "Course-related Reading in English" correlates weakly only with Background Knowledge (rho = .141** at p = .008). These results suggest that students' reading habits could be considered one of the possible predictors of Reading Comprehension. What appears to be standing out among these results is the fact that Course-related reading does not correlate with Reading Comprehension, whereas General reading does. Moreover, Background Knowledge correlates more strongly with General Reading than with Course-related Reading, which is also somewhat unexpected.

Gender data was analysed with Mann-Whitney test and showed a significant difference in means between male and female students with regard to their "General Reading in English" and their "Course-related Reading in English". The result indicated that male students read more on general topics and females slightly more on course-related topics (see Table 5.41). This suggests that there is some difference in the reading focus between female and male students.

7.10 Model of reading comprehension

Combining results from the testing and think-aloud studies provided evidence that building a coherent model of text comprehension involves readers' ability to process and integrate information at the sentence, paragraph, and text level with the knowledge they possess. The think-aloud findings suggest that low background knowledge readers took more of a local approach to reading. They attended to single individual words 100% more often than high background knowledge readers as indicated by the totals of single word translations, either correct, approximate, or wrong. Low knowledge readers tended to process the text word by word, they had difficulties in building connections between the parts of the text, they especially had problems to integrate new information or find the main idea. They may have understood parts of the text, but struggled to integrate them into a meaningful whole. Some readers were aware of their inability to see the bigger picture and build a coherent text model, so they expressed their frustration in their monitoring comments: "I understand every word, but I don't understand everything together" (TA11).

Overall, these findings corroborate the results of studies that found support for the position that low background knowledge readers focus more on the surface code and literal meaning (Chang, 2006, p. 177), word and sentence level rather than text level (Yamashita, 2002; Nalliveettil, 2014), they attend to local clues rather than global (Bengeleil & Paribakht, 2004). Stahl et al. (1991) make a distinction between comprehension of microstructure and macrostructure. They suggest that vocabulary difficulty affects more the tasks related to microstructure and individual propositions, whereas domain knowledge seems to affect more the comprehension tasks that involve macrostructure or the representation of gist and most important text information.

Readers' local approach to reading appeared to be linked to their clinging to the text model of interpretation, also referred to as reading close to the text. This has been observed as typical of low knowledge readers (Grabe, 2009, p. 11).

> "… students with low-domain, or background, knowledge read with the purpose of gaining comprehension from the close reading of a text. Students with high-domain knowledge read to build a more elaborative knowledge base and draw more selectively from the text."

According to Van Dijk and Kintsch (1983, p. 51), readers construct a situation model of comprehension by linking the textbase to their knowledge through inferences functioning as the links in the process. In this study, high background knowledge readers elaborated, inferred and evaluated more than low background knowledge readers, thus creating their situation model of comprehension. However, there was great individual variation among readers with respect to how much they elaborated, inferred or evaluated, a finding similar to Bråten and Strømsø (2003). This result appears to indicate individual differences in depth of readers' background knowledge as well as individual differences in readers' personality traits.

Think-aloud data analysis provided evidence for the conclusion that readers in this study showed variable level of flexibility in the reading comprehension process. Some readers were able to read selectively, go back and forth in the text to check information, and adjust their understanding. They may have skipped parts of the text they found unimportant or too difficult, and they came back to them later and changed their understanding. A flexible approach to reading appeared to be closely related to readers' ability to tolerate the unknown and accept ambiguity. This means that readers were able to carry on reading with comprehension despite failing to understand every single word or sentence. It is also important to note that low knowledge readers seemed to be more fixated on what they did not know and they more frequently reported on what they did not understand. A key difficulty for readers was parsing or working out the grammatical structure of sentences, sentence elements and parts of speech. This was manifest in readers' problems to understand the relationships expressed in sentences and it crucially affected and impaired their comprehension. The parsing problem may have been related to the density of difficult words that some readers have commented on in their think-alouds. Finally, despite having learnt English for over seven years, some readers had difficulties with lower-level reading skills, such as the ability to decode the words correctly. They made decoding mistakes that affected their comprehension. To illustrate, instead of 'imagination' a reader would read 'immigration', instead of 'correlated' 'corrected', instead of 'idea' 'ideal', or instead of 'financial theorists' 'financial terrorists', as shown in the following example:

> Text: For financial theorists who model the value …
>
> Think-aloud: *Financial terrorists – probably they are getting money in illegal ways* (TA 2 BRR)

Even at an advanced level reading, lower-level processing skills including decoding can be a problem for some readers. Automatizing lower-level processing affects higher-level processes, hence it ought to be considered in reading instruction (Grabe, 2009; Paran, 1996).

7.11 Background knowledge and vocabulary

The results of the testing and the think-aloud studies have provided evidence for the position that readers' depth and breadth of background knowledge, including vocabulary impacts their processing of subject-specific texts. Readers in both studies were very much aware of that and when they gave comments on the difficulties they encountered in reading subject-specific texts, among the most frequently given reasons were vocabulary and background knowledge. However, there is no clear line between background knowledge and vocabulary. Ash and Bauman (2017, p. 379) pointed out that

> "Readers' general conceptual knowledge promotes or causes reading comprehension, not word knowledge per se. Instead, vocabulary knowledge is indicative of a reader's broader knowledge base about a topic and the words used to describe it."

Several other researchers have noted the inseparability of knowledge and vocabulary. They emphasized that background knowledge is articulated through vocabulary (Fisher & Frey, 2009) or that it manifests itself as vocabulary knowledge (Marzano, 2004). In a similar vein, Rydland et al. (2012, p. 467) state:

> "It is difficult to disentangle the construct of prior topic knowledge from the construct of vocabulary."

As there are links between both constructs, they may share variance in comprehension. Comparing vocabulary breadth (i.e. number of words) and depth (i.e. richness of semantic representation), Rydland et al. (2012) conclude that among different vocabulary measures used, vocabulary depth explained the largest variance in reading comprehension. In their comprehensive reviews of vocabulary learning and teaching, Nation (2001) and Schmitt (2010) underline this inseparable link between knowledge and vocabulary by saying that learning a discipline involves learning its specialized technical vocabulary. Webb and Nation (2017, p. 15) also say:

> "Technical words represent specialized knowledge that is essential to learning a particular topic …We usually learn technical vocabulary as we learn about a topic."

While giving a similar position about vocabulary-knowledge relationship, Bailin and Grafstein (2016) stress that readability of a text must be assessed in relation to readers' background knowledge because readers vary in their background knowledge.

Overall, the quoted studies suggest a strong relationship between vocabulary and reader's knowledge, whereby vocabulary without conceptual knowledge does not facilitate comprehension. This was manifest as a recurring pattern in this study. In the think-aloud study, readers may have found a word familiar, but they did not know the meaning and concept behind the word, which impaired their understanding. An important implication for reading instruction can be drawn here: to facilitate effective comprehension, words in readers' lexicon should be paired up with broader conceptual knowledge.

In their semantic processing, readers select content appropriate meaning (Koda, 2005). They match their knowledge to the textbase and if there is congruence between the two, the comprehension is effective. This is particularly evident in cases when texts are challenging due to lexical, semantic and discourse complexities, as well as because of the amount of implied information. Grabe (2009) claims that reading expository texts for study purposes significantly affects processing as readers use more inferences, paraphrases, and recall more information. He also notes that reading for study purposes requires that readers draw more on their background knowledge; they are very likely to activate more information to aid their text interpretation, elaboration on it and critical evaluation. Similarly, Cervetti and Wright (2020) observe that reading expository texts induces readers to make more prior knowledge elaborations than readers of narrative texts as they need to do more to integrate text and recall it.

An expository text can be dense and have a high proportion of content words to function words; it can contain concepts that are unexplained or new to the reader. So, readers need to use their knowledge to fill the gaps in the text and to use context to make connections between ideas. All of these are challenges for the readers and they deal with them in strategic ways. In this study, the following processing patterns were observed when readers dealt with difficulties.

1. Adjusting understanding

   When faced with a difficult or unknown word or sentence, some readers changed their interpretation as they read further and amended their understanding of the text. For example, some readers first interpreted 'royalty' as 'the royal family' but after reading on they realized that this meaning did not fit the context, so they changed the interpretation to 'some type of payment', which corresponds to the financial context.

2. Clinging to wrong understanding

   Some readers clung to incorrect understanding even if a particular expression was explained in the subsequent text. For example, the term 'security' meaning a financial instrument was interpreted incorrectly to the context as 'insurance', 'personal safety', 'protection' 'guarding' or 'intelligence' although the meaning did not make much sense in the context. Similarly, 'to securitize' meaning pooling financial instruments was incorrectly interpreted as 'to protect', 'to guard', 'to safeguard'. Low knowledge readers did not pick on inappropriacy in the given context and they nevertheless stuck to this distortion when the same word appeared again later in the text.

3. Inconsistencies in understanding

   Some readers were inconsistent in their think-alouds. They randomly changed and switched their translations of the same expression as the text progressed. Inconsistencies as well as clinging to wrong interpretations can be illustrated by extracts form think-alouds given by TA 14 from LBK group:

Text: … a bond like **security** that the most conservative of investors might buy …
Think-aloud: *This was **safety** that can be bought only by the richest.* (TA 14)

Text: … banks **securitizing** loans …
Think-aloud: *The **safety** of the banking system*… (TA 14)

Text: This makes **securities** difficult to value and therefore hard to trade.
Think-aloud: ***Intelligence** services get certain value, their services become valuable*. (TA 14)

Text: … **securitization** …
Think-aloud: *They have their own* **bouncers, security people**. *They don't need to hire them.* (TA 14)

Text: Indeed such **securitizations** have not become as fashionable as many expected.
Think-aloud: Some sort of **insurance** … (TA 14)

4. Failure to understand basic vocabulary
   It was surprising that readers were unfamiliar with some vocabulary that is not highly specific like: 'board' or 'compensation'.

To interpret the results related to vocabulary, I used Nation's (2001, p. 198) classification of specialized vocabulary into three categories. Category 1 technical *words* or *terminological words* appear rarely if at all outside the field (examples in this study were: 'equity', 'exogeneous', 'endogenous', 'imprimatur'). Nation links category 1 technical words to one's knowledge of disciplines by stressing that if you know a discipline, you know its terminology. The findings of this study provide evidence to support Nation's claim. Low knowledge readers found terminological words especially difficult. Because terminological words form an integral part of background knowledge of disciplines, the low background knowledge readers are disadvantaged in their reading comprehension.

The majority of readers in this study had considerable difficulties with category 2 technical words (Nation, 2001, p.199). These words are formally like high-frequency words but they also have specialized meanings and are used in and out of the field with a different meaning (examples in this study: 'bond', 'security', 'default', 'royalty', 'options'). These words proved to be especially challenging for low background knowledge students, as they would tend to use the non-specialized meaning of the word despite realizing that it was not appropriate in the context. For example, when thinking aloud about ‘bonds’, they talked about ‘connections between people’ rather than ‘special types of securities’ in finance; to refer to ‘default’, they used the meaning from IT rather than ‘non-payment’ from financial context. The word ‘security’ was a recurring problem just for the fact that on surface it appeared familiar, and was interpreted as ‘safety,’ ‘insurance,’ ‘guarding’ or even ‘bouncers’ rather than ‘a financial instrument with special characteristics’, the meaning that actually corresponds to the context. The evidence from this study suggests that Category 2 technical words are exceptionally challenging for the readers due to their different meaning in a specialized and general context. Similarly, in their discussion of the readability of finance texts, Bailin and Grafstein (2016), stress that specialized financial vocabulary poses serious problems for non-specialists, including words with non-financial meanings due to their semantic ambiguity.

Category 3 technical words (Nation 2001, p. 199) seemed to be slightly less of a challenge for the readers in this study. These words are used both inside and outside the field and their specialized meaning can be accessed through non-specialized meaning (examples in this study: ‘downgrade’, ‘share’). For instance, readers correctly drew a conclusion that if a share was ‘downgraded’, this means that it was worth less. Similarly, ‘shares’ were correctly interpreted as parts of ownership.

7.12 Chapter summary

This chapter presented a discussion of the answers to the research questions based on the results of the testing and think-aloud studies. The results were contextualized and considered in comparison to the existing research and theory in L2 reading. The focus was an evaluation of the role of Background Knowledge in Reading Comprehension of subject-specific texts in contrast to the role of Grammar Knowledge and Self-reported Familiarity in the same process. Whereas Background Knowledge and Grammar Knowledge were both found to be predictors of Reading Comprehension, this was not the case with Self-reported Familiarity. Hence, a distinction was drawn between the two variables, suggesting a need for testing knowledge rather than self-reporting or assuming it. The relationship between Background Knowledge and Grammar Knowledge was discussed with regard to the results of testing for the threshold and compensation effects. The discussion of the results obtained in the testing study was paired up with the findings from the think-aloud study and readers' comments in the post-reading questionnaire. Typical reading patterns and strategies used by readers with high or low levels of Background Knowledge were compared. The chapter ended with a presentation of the emerging L2 reading model and a discussion of the relationship between background knowledge and vocabulary.

CHAPTER 8
CONCLUSION

8.1 Chapter aim and overview

This chapter gives a final overview of this study. It provides a summary that begins with research aims, objectives, and questions followed by an outline of the methodology and main findings. The chapter then considers the study's contributions and discusses the implications both for L2 reading instruction and reading research. Finally, it turns to the limitations of the study and offers directions for future research.

8.2 Summary of the study

Current academic settings typically require L2 reading of texts in various disciplines, and effective L2 (especially EFL) reading comprehension is often a marker of academic success. What makes L2 reading in disciplines different from general reading is the fact that it relies on readers' L2 knowledge and discipline-related background knowledge (Cervetti & Wright, 2020; Uso-Juan, 2006). This relationship has been examined in theoretical and empirical studies and the results have been mixed. This study set out to investigate the relationship between background knowledge and grammar knowledge in L2 reading of subject-specific texts, by using two operationalizations of background knowledge: tested background knowledge and self-reported familiarity. The first aim was to determine the strength of the relationship among predictor variables: Background Knowledge, Self-reported Knowledge, Grammar Knowledge, and the criterion variable Reading Comprehension. Apart from exploring the relationship and influence of variables, the second aim was to investigate the processing patterns and strategies of readers with high and low background knowledge. The third aim was to link the quantitative and qualitative data obtained in the testing and think-aloud studies to get an overall picture of the relationships among the studied variables.

Based on the research aims, five specific research objectives were defined. First, to identify and compare the contributions of Background Knowledge in L2 Reading

Comprehension with regard to its strength and interaction with L2 Grammar Knowledge; second, to test the existence of any L2 threshold necessary for the effect of Background Knowledge; third, to test the existence of compensation between Background Knowledge and L2 Grammar Knowledge; fourth, to compare the contribution of tested and perceived background knowledge in L2 reading by using two different operationalizations of background knowledge: a test of text-related background knowledge on one hand and readers' self-reports of their background knowledge; and fifth, to compare patterns and strategies used by readers with different levels of background knowledge.

The research aims and objectives led to the formulation of eight research questions and subquestions (see Section 2.6), which were addressed by conducting the testing and think-aloud studies. The main variables were operationalized as follows:

1) Background Knowledge as a score on a test of background knowledge of finance related to three reading texts;
2) Grammar Knowledge as a score on a test of grammar;
3) Reading Comprehension as a score on a reading comprehension test based on three subject specific-texts;
4) Self-reported Familiarity as readers' rating of topic familiarity on a 7-point scale.

The texts were selected by finance experts according to five criteria (i.e. specificity, appropriacy, familiarity, new knowledge, and explanation). Finance experts were also involved in item writing for both the reading comprehension test and the background knowledge test. The latter tested the knowledge finance experts deemed vital for comprehension of the selected three texts. All instruments were piloted and validated with the assistance of language experts and finance experts. Altogether 404 students took part in the study, 22 in the piloting study and 382 in the main study, out of which 358 were engaged in the testing study and 24 in the think-aloud study. The dataset obtained after the administration of all tests and think-aloud procedure comprised the results of readers'

1) grammar tests,
2) background knowledge tests,
3) reading comprehension tests based on three subject-specific texts,

4) post-reading questionnaires, and
5) think-aloud verbal protocols.

In the testing study all tests were scored, and the scoring was validated with the assistance of finance experts.

To answer the research questions, the datasets obtained from both studies were analysed. The testing study data were analysed using several statistical methods: descriptive statistics, bivariate correlation analysis, multivariate analyses (multiple regression), and non-parametric tests (Kruskal-Wallis H test, Mann-Whitney U test). In the think-aloud study, verbal protocols were coded using a data-driven coding scheme, and the frequencies of coding categories were tallied. Next, the results were linked to readers' reading comprehension scores and compared between two groups of readers with high or low background knowledge.

The research design and research questions of this study are based on the literature review of background knowledge studies and open questions arising from it. This study uses triangulation with regard to: 1) theoretical and empirical approaches consulted, 2) data types generated, and 3) research methods used. As regards the consulted approaches, the study reviews relevant theories and empirical studies in L2 reading comprehension as well as think-aloud processing. Second, the study combines two sources of data: quantitative, such as test scores that are subjected to a range of statistical analyses, and qualitative data, which are derived from the coding of think-aloud protocols. Finally, this is a mixed-methods study that integrates two research methodologies: the testing methodology on the one hand and the think-aloud methodology on the other. Thus, it combines the product and process methodological approaches to answering eight research questions. The main findings that answered them are summarized next.

RQs 1 and 2 Contribution of Background Knowledge, Grammar Knowledge, and Self-reported Familiarity to L2 Reading Comprehension

To answer this question, several steps were taken by running different types of statistical analyses, from bivariate correlation to multiple regression. The first results of bivariate analysis showed that the three independent variables (i.e. Background Knowledge, Grammar Knowledge, Self-reported Familiarity) were statistically significant and positively correlated with Reading Comprehension. The correlation of Reading Comprehension with Background Knowledge ($rho_{BK} = .471^{**}$ at $p < .001$) was slightly stronger than with Grammar Knowledge ($rho_{GK} = .412^{**}$ at $p < .001$), whereas correlation with Self-reported Familiarity was the lowest ($rho_{SRF} = .162^{**}$ at $p = .002$). As correlation only reveals the strength of the relationship but not causality or the predictive value of a variable, a further step was required and the data were subjected to multiple regression analyses, which provided complete answers about the predictors of Reading Comprehension. With 29.5% of variance in Reading Comprehension accounted for by the variables, the results showed that only Background Knowledge (standardized $\beta = .343$ at $p < .001$) and Grammar Knowledge (standardized $\beta = .312$ at $p < .001$) are statistically significant predictors of Reading Comprehension at a medium level, whereas Self-reported Familiarity did not turn out to be a statistically significant predictor. The explained variance in other studies was similar (Clapham, 1996) or higher (Uso-Juan, 2006). The results in this study are in accordance with findings of those studies that found stronger support for the role of both background knowledge than language knowledge in L2 reading comprehension (Lin & Chern, 2014; Hwang, 2019; Hwang & Duke, 2020). In this study, the predictive value of Background Knowledge was found to be significant at medium strength and slightly stronger than Grammar Knowledge. This may be due to several factors. First, in this study Background Knowledge was tested and not assumed. Second, reading comprehension texts were selected by finance experts against pre-set criteria (i.e. specificity, appropriacy, familiarity, new knowledge, and explanation) in order to make it possible for readers' background knowledge to be shown, as implied by Grabe (2001, p. 234):

> "…when reading material does not make strong demands on specialist topical knowledge, the supportive effects of topical knowledge on comprehension decrease."

Based on the results of correlation and regression analyses, it can be concluded that when reading subject-specific texts, readers tend to draw on background knowledge and grammar knowledge to construct their mental representation of the text (i.e. the situation model, Kintsch, 1988); they both have a facilitative role in comprehension, the role of background knowledge is slightly stronger.

An important additional insight was obtained when seeking the answers to RQs 1 and 2. A comparison of Background Knowledge and Self-reported Familiarity in this study revealed that the relationship between Background Knowledge and Reading Comprehension is much stronger than the relationship between Self-reported Familiarity and Reading Comprehension, although both are statistically significant. However, when multiple regression was run, the results showed that Self-reported Familiarity was not a statistically significant predictor and thus cannot be compared to Background Knowledge. This highlights the distinction between the two operationalizations of background knowledge used in this study in favour of tested background knowledge. It calls into question the use of Self-reported Familiarity and its validity. It appears that when researching discipline-related background knowledge in L2 reading, Self-reported Familiarity cannot be taken as a valid measure of discipline-related knowledge because it is not statistically significant. In addition, this distinction was supported by the finding of bivariate analysis, which shows a weak correlation between Background Knowledge and Self-reported Familiarity (rho = .196** at $p < .001$), meaning that both variables pull in the same direction, but differ in terms of strength. Despite its popularity as knowledge operationalization in L2 reading, the use of self-reported familiarity has been criticized (Cervetti & Wright, 2020; Chou, 2011) and more recent studies exploring background knowledge have not used self-reports or other assumption-based operationalizations of background knowledge, but they have rather opted for tests of background knowledge (Uso-Juan, 2006; Erçetin, 2010; Rydland et al., 2012; Kelly, 2014; Lin & Chern, 2014; Karimi; 2017; Hwang, 2019; Hwang & Duke, 2020; Song & Reynolds, 2022).

The relationship between Background Knowledge and Self-reported Familiarity is not only related to the type of measurement, but also to what is being measured. Familiarity is a concept that is shallower, whilst background knowledge is deeper because it contains the knowledge of concepts and relationships (Chou, 2011; Taboada & Guthrie, 2006). In discipline-related background knowledge, the depth refers in particular to technical vocabulary, which is learnt when one learns a particular discipline, therefore it needs to be taken as an indicator of one's background knowledge (Nation, 2001) Both the type of measurement and the construct measured need to be considered when the background knowledge variable is explored. The findings in this study suggest that there is a distinction between Self-reported

Familiarity and Background Knowledge that merits consideration when designing reading studies and operationalization of variables.

The influence of predictor variables was also tested for the individual texts "Star Struck" (SS), "Taking Stock" (TS), and "Bonds that Rock and Roll" (BRR), and it confirmed the results obtained for all texts together. The influence of Background Knowledge on Reading Comprehension was similar for each individual text and all texts together. However, the influence of Self-reported Familiarity was different because the multilevel regression analysis showed it had some influence on reading comprehension of one text (BRR), but not on all texts together. This finding also confirms the limited role of Self-reported Familiarity in Reading Comprehension. While Self-reported Familiarity may have a limited role as a predictor of Reading Comprehension of individual texts, this cannot be claimed for all texts together.

RQs 3 & 4 The threshold effect

The interaction between background knowledge and language knowledge has been at the heart of discussion in L2 reading studies. Overall, it has been reported that the deployment of background knowledge tends to depend on the level of readers' language knowledge, which led to hypotheses and conclusions about lower or upper thresholds (Clapham, 1996; Uso-Juan, 2006; Ridgway, 1997; Karimi, 2016). However, the empirical evidence in this study does not provide support for thresholds. This conclusion is based on a multiple regression analysis that involved readers divided into nine groups according to the combinations of their level of Background Knowledge and Grammar Knowledge (i.e. low, medium, high). The effect of Background Knowledge was detected at all three levels of Grammar Knowledge. This result appears to contradict Ridgway's (1997, p.161) claim:

> "While background knowledge is always at work, the effect is not always detectable. It appears that a reader must be at a certain level with relation to the text for the effect to be observable."

It needs to be acknowledged, however, that the background knowledge effect in this study was slightly weaker with readers that had a high level of Grammar Knowledge; however, the difference was not so large that it could be interpreted as an upper threshold. This finding does not corroborate Clapham (1996) and Uso-Juan (2006), who found a stronger impact of Background Knowledge for low and intermediate proficiency readers, and both confirmed the lower threshold. In contrast, this study's finding is in accord with Krekeler (2006), who similarly did not find support for thresholds. What especially resonates with this study is his suggestion that if there are thresholds, they are fuzzy and they crucially depend on subject specificity, an observation also shared by Grabe (2009) and Urquhart & Weir (1998).

RQ 5 & 6 The compensation effect

This study provided empirical evidence to confirm the compensation effect by comparing nine groups of readers with varying combinations of background knowledge and grammar knowledge. Initial comparison of groups' reading comprehension scores showed that high knowledge readers performed better on the reading comprehension test than low knowledge readers. Next, the compensation effect was statistically tested by running the Kruskal-Wallis H test, and the results showed that both "compensating groups" (Low GK-High BK group and High GK-Low BK group) outperformed other groups apart from High GK-High BK group. The results confirmed the compensation effect both with regard to Background Knowledge and Grammar Knowledge, suggesting that when faced with comprehension difficulties, readers were able to leverage the knowledge resources at hand and make up for the lack of one resource with another. Several other studies have reported the compensation effect (Koh, 1985; Chen & Donin, 1997; Al-Shumaimeri, 2006; Uso-Juan, 2006; Hwang & Duke, 2020). Uso-Juan (2006) confirmed compensation for both background knowledge and L2 proficiency, which is in line with the findings of this study. However, the effect of background knowledge was found to depend on several other factors: first, operationalization and measurement of background knowledge; second, text features, especially text specificity; and third, the statistical analyses used (Cervetti & Wright, 2020). Although deemed important, it has not yet been agreed what makes a text subject specific (Uso-Juan, 2006). Subject experts may disagree about it, but if specificity is not controlled for, a background knowledge effect is not likely to be confirmed (Clapham, 1996). According to Nation (2001), it includes field-specific terminology, whereas Douglas (2000, p. 33) contends that subject specificity critically depends on the amount of unexplained subject-specific concepts and contextembedded information, and therefore needs to be determined by experts in the field. Douglas (2001, pp. 46–50) concludes that language knowledge is an inseparable part of field-specific background knowledge and the two concepts are inextricably intertwined. Readers use background knowledge and language knowledge in an additive and compensatory way (Horiba & Fukaya, 2015). This suggestion was confirmed by the compensatory effect found in this study.

RQ 7 Processing patterns and strategies of high and low background knowledge readers

Having analysed the verbal protocol data and combining them with reading comprehension scores, it was possible to determine and compare reading patterns and strategies used by readers with high and low background knowledge. The comparison between the two groups of readers showed differences with respect to the types and the frequency of strategies used. Although both groups used the same strategies, there were differences in the frequency of use. The high background knowledge group used more cognitive strategies: elaboration, inferences and evaluations, and correct paraphrases; whereas the low background knowledge group had a more local-level approach by focusing more on individual words, phrases, and sentences, which may have been a result of their lack of vocabulary. They reported comprehension problems and had difficulty tolerating the unknown or irrelevant in order to focus on more important information and seeing the bigger picture. Overall, the comparison of results between the two groups with high and low background knowledge tends to suggest differences in semantic and pragmatic processing at the local and global level.

To shed more light on readers' views of the reading process, the post-reading questionnaire contained questions about the sources of difficulty, interest, and familiarity. Bivariate analysis revealed that all variables were significantly correlated with Reading Comprehension, but the strongest was the negative correlation with Difficulty, suggesting that readers' assessment of difficulty was indicative of their test scores. Readers reported that the major source of difficulty was unknown specialist vocabulary. Two other medium correlations among variables emerged, indicating that Self-reported Familiarity correlated with lower Difficulty and Interest with higher Self-reported Familiarity.

8.3 Contribution of the study

This study adds to research in L2 reading in the following four ways.

First, the study empirically shows that there is an important distinction between tested Background Knowledge and Self-reported Familiarity with regard to their effect on L2 Reading Comprehension of subject-specific texts. While tested Background Knowledge had a statistically significant impact of medium strength on L2 Reading Comprehension, Self-reported Familiarity did not have any impact. Both measures have been used in a number of previous studies and have often been treated as interchangeable, despite some critique noting this as an inconsistency of L2 reading studies with regard to measurement and operationalization (Cervetti & Wright, 2020; Shin, 2010). In addition, it has been pointed out that there is a difference between familiarity and knowledge with regard to the depth of the concept (Chou, 2011). While familiarity refers to one's surface acquaintance with a topic, knowledge entails a deeper understanding of the concepts and relationships involved in a particular topic, area, or field. This suggests that familiarity may share a portion of background knowledge construct, but it cannot be treated as interchangeable. This was confirmed by the result of regression analysis revealing that Background Knowledge and Self-reported Familiarity share a weak yet statistically significant relationship (rho = .196** at $p < .001$).

This study responded to Uso-Juan (2006), who raised the question whether self-reported familiarity can substitute for testing knowledge. To empirically explore the difference between Background Knowledge and Self-reported Familiarity, this study used both operationalizations, which has not been done in one single study before. The results of multiple regression have shown that only tested Background Knowledge was a statistically significant medium predictor of Reading Comprehension of subject-specific texts, whereas Self-reported Knowledge was not. Hence, the study tends to suggest that there is a difference between the effect of tested background knowledge and self-reported familiarity in L2 reading comprehension. This distinction may be especially important in L2 studies involving discipline-related knowledge. This study adds empirical grounds for a critical consideration of the measurement of knowledge used in L2 reading studies. It contributes to rethinking of the measurement and operationalizations in the design of L2 reading studies that, like this one, involve discipline-related background knowledge.

Second, this study provides insights into the role of Background Knowledge and Grammar Knowledge in L2 reading, by specifying the relative contributions of variables and their predictive value in L2 Reading Comprehension. It confirms the significant facilitative role of both variables based on the results of bivariate correlation and multiple regression analyses. It also highlights the interaction between Background Knowledge and Grammar Knowledge, which was manifest in the compensation effect as readers with deficiencies in one resource could make up for it with better knowledge of the other. Overall, the empirical evidence in this study confirms that having text-relevant background knowledge of the discipline benefits comprehension and works in tandem with grammar knowledge, thus supporting Horiba and Fukaya's (2015) view that background knowledge and L2 knowledge act in an additive and compensatory way.

The third contribution of the study is a methodological one. It responded to the need for more mixed-method studies identified by several researchers (Horiba & Fukaya, 2015; McNeil, 2012; Uso-Juan, 2006) by combining the testing study and think-aloud study to render both quantitative and qualitative datasets. In this way, this study not only reveals the findings about the importance and strength of the variables in L2 reading, but also suggests how readers with different levels of knowledge read subject-specific texts. The use of mixed-methods in this study offers a better understanding of the research problem, which is argued to be a key benefit of this approach (Creswell, 2014; Loewen & Plonsky, 2016).

Finally, in accordance with Douglas (2001), the study involved finance experts at all stages of instrument design, test scoring, verbal protocol coding, and validation. This was felt to be of particular importance when dealing with subject-specific texts and discipline-related knowledge. Text selection reflected the authentic reading that the participants in the study do as part of their coursework and the final selection of texts was made by using a set of five criteria including: appropriacy, subject-specificity, familiarity, new information, and the amount of explanation. The background knowledge test was specifically related to the three texts by testing information and concepts that financial experts considered to be essential for text comprehension.

8.4 Implications of this study

The implications of the findings of this study can be related to L2 reading instruction and future reading research.

8.4.1 Implications for L2 reading instruction

The findings of this study, both the predictive values of variables and comprehension patterns used by high and low background knowledge readers, provided evidence to discuss possible implications for L2 reading instruction.

First, this study has empirically confirmed the role that Background Knowledge plays in facilitating L2 Reading Comprehension. This finding resonates with metaphoric observations that background knowledge acts as a driver and filter of comprehension and learning (Urquhart & Weir, 1999), as well as a scaffolding of information in the reading process (McNamara, 2007). However, its facilitative effect can only take place if knowledge is activated (Cervetti & Wright, 2020). It is not expected that L2 reading instructors would explicitly teach background knowledge, but rather that they would help activate readers' existing knowledge either by using various pre-reading activities (Brantmeier, 2003) or by training of strategies, especially lexical inferencing (Elbro & Buch-Iversen, 2013) or mapping strategies (Chou, 2008).

Second, in order to train readers to use their discipline-related background knowledge and develop learners' disciplinary literacy, L2 reading instruction should expose readers to ample quantities and types of discipline-specific texts (Duke et al., 2011). Instruction ought to provide opportunities for the readers to see how specialist discipline knowledge is communicated through vocabulary, structure, and structural patterns of expository texts. Readers need to be trained how to use specialized reading skills and practices required for disciplinary reading (Shanahan, 2017). The goal is to develop literacy that corresponds to the discipline, so that readers can approach texts as practitioners in the discipline. Duke et al. (2011) outline essential elements of effective reading comprehension instruction, among which they list building disciplinary and world knowledge, providing exposure to a volume and range of texts, providing motivating texts, teaching comprehension strategies, teaching text structures, building vocabulary and language knowledge, and integrating reading and writing.

Third, the findings of this study, especially readers' comments about the sources of difficulty in texts and results of think-aloud study, have reflected the inextricable link between disciplinary knowledge and vocabulary. Hence, this study suggests that special attention in L2 reading instruction needs to be paid to the teaching of the three categories of technical vocabulary (Nation, 2001). The results of the think-aloud study indicated that especially category 2 technical words pose a major problem because of their multiple meaning in different contexts. With respect to these words, there is need for vocabulary-focused activities that relate technical uses of a word to its core meaning (Nation, 2008). This would make it possible for the learners to become aware of the areas of overlap and the areas of difference between the technical and non-technical uses of the word.

Finally, leveraging background knowledge in the reading process concerns higher-level processes, however, these processes depend on the automaticity of lower-level skills. The think-aloud study showed that some readers still have difficulties at that level. Therefore, a corollary of this would be that L2 reading instruction needs to provide opportunities for the targeted development of lower-level skills, so that readers can reach decoding automaticity (Jeon & Yamashita, 2020). The findings of this study offer evidence to support Paran's position (1996) that good readers have fluent and automatized lower-level skills, therefore L2 teachers should assist readers to develop automaticity through word-recognition training to become good decoders.

8.4.2 Implications for L2 reading research

The key implication for research based on this study's findings is the distinction between tested Background Knowledge and Self-reported Familiarity. Only Background Knowledge affected Reading Comprehension and was found to be its predictor. This was not the case with Self-reported Familiarity. Based on this multiple regression finding, the study suggests rethinking of the operationalization and measurement of background knowledge in L2 reading research, especially when it involves discipline-related knowledge.

The second implication that may also refer to reading instruction concerns the central role of vocabulary. It is based on the evidence from the think-aloud study and readers' attributions of sources of difficulty in the post-reading questionnaire. In this study, specialist technical vocabulary appeared to be at the heart of the problem due to its convergence with background knowledge and text subject-specificity. Readers showed their awareness of the key importance of vocabulary as a source of difficulty in reading comprehension. The findings of this study support for Webb & Nation's (2017) claim that learning a topic means learning technical vocabulary. This suggests that future L2 reading research involving background knowledge would need to systematically involve the vocabulary element.

8.5 Limitations and future research

Despite carefully planning the research design and principled undertaking of the experiments and analyses in this study, there are several limitations that need to be acknowledged.

To begin with, participants were Slovene university students of business and economics, with L1 Slovene, so their L1 and background could have had an effect that may restrict the generalizability of findings and suggest reasons for more research involving participants with other L2's and backgrounds. It would be beneficial to conduct research with participants with other L1 and L2 and different ages.

Second, the explained variance in the regression analyses in this study ranged from 29.5% to 30%, which means that a large portion of variance remained unexplained. In comparison to other studies, this figure is similar to Clapham (1996), McNeil (2012), and lower from Uso-Juan (2006).

Third, to maintain the authenticity not only in text selection, but also at the task level, subject experts who were also the participants' subject teachers were asked to do item-writing for the reading comprehension test without any interference from the researcher. The downside of this was that test items were not controlled for the degree of literal and inferential comprehension they required and with regard to the mix of textually explicit or implicit question types. This may have affected the results and ought to be screened for in future research.

Fourth, the distribution of test scores in this study was specific, with the grammar knowledge test having relatively higher scores in contrast to the relatively lower scores in the background knowledge test. The lower scores in the background knowledge test could be due to the fact that the original brief to the subject experts in the selection of texts was to select texts that would make it possible for the readers to demonstrate in their reading comprehension whether they have text-relevant background knowledge. This may have directed subject experts towards the choice of more demanding texts. It could also be linked to the differences in scores among the three texts in the main study, in which readers found the BRR text more difficult and more unfamiliar than other texts. Readers' observations were corroborated by the differences regression results for individual texts. In the text selection process, Gunning Fog and Flesh-Reading Ease readability formulas were used as orientation of text difficulty. It needs to be acknowledged that these tests have mostly been used to evaluate text difficulty in L1 studies and it has been noted that their use in L2 may not be entirely straightforward (Eidswick, 2010).

Fifth, as regards the use of L1 and L2 in the study, it needs to be mentioned that readers in think-aloud sessions were given the option to either use their L1 or L2. However, in the test of reading comprehension and background knowledge test only L2 was used, which may have, to some extent, affected the results. In the future, it would be better to give readers the option of using L1 in the test of background knowledge.

Sixth, due to the size of the test battery, it was essential to control the length of administration. I tried to see to that by piloting the instruments and subsequently adjusting them so that they would be more practical and less time-consuming while still maintaining reliability. This was particularly important in the think-aloud study, in which the results may have been somewhat affected by the fatigue factor. I tried to partly avoid that by rotating the sequence of texts in the battery for each student in the think-aloud sessions. A test sequence rotation was also used in the testing study. Despite the precautions, the results could have still been affected by the fatigue factor.

Finally, after the think-aloud session with every reader, I had a conversation about how they had read the texts, their topic familiarity and knowledge of concepts, as well as how difficult and interesting they found the texts. I recorded the conversations and took notes, but these were not structured interviews that would be systematically analysed. With hindsight, structured interviews could have provided better understanding of readers' perceptions of the reading process.

LIST OF REFERENCES

Adams, S. J. (1982). Scripts and the recognition of unfamiliar vocabulary: Enhancing second language reading skills. *Modern Language Journal*, 66 (2), 155-159.

Afflerbach, P. (1990). The influence of prior knowledge on expert readers' main idea construction strategies. *Reading Research Quarterly*, 25 (1), 31-46.

Alderson, C. (2000). *Assessing Reading.* Cambridge: Cambridge University Press.

Alderson, J. C. & Urquhart, A. H. (1983). The effect of student background discipline on comprehension: A pilot study. In A. Hughes & D. Porter (Eds.), *Current Developments in Language Testing.* London: Academic Press.

Alderson, J. C. & Urquhart, A. H. (1984). Postscript on Hosenfeld. *Reading in a foreign language*, 245-249.

Alderson, J. C. & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing,* 2(2), 192-204.

Alderson, J. C. & Urquhart, A. H. (Eds.) (1984). *Reading in a foreign language*. London: Longman.

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp.1-24). Harlow: Longman.

Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Al-Shumaimeri, Y. (2006). The effects of content familiarity and language ability on reading comprehension performance of low- and high-ability Saudi tertiary students studying English as a foreign language. *Education Science and Islamic Studies,* 18(2). 1-19.

Alvermann, D. E., Smith, L. C. & Readence J. E. (1985). Prior knowledge activation and the comprehension of the compatible and incompatible text. *Reading Research Quarterly,* 20(4), 420-436.

Amadieu, F., Salmeron, L., Cegarra, J., Paubel P. V., Lemarié, J. & Chevalier, A. (2015). Learning from concept mapping and hypertext: An eye tracking study. *Educational Technology & Society* 18(4), 100-112.

Anderson, R. C. & Pearson, P. D. (1984). *A schema-theoretic view of basic processes in reading comprehension*. Technical Report No. 306.

Anvari, S. & Farvardin M. T. (2016). Revisiting lexical inferencing strategies in L2 reading: A comparison of successful and less successful EFL inferencers. *The Reading Matrix: An International Online Journal*, 16(1), 63-77.

Ash, G. E., & Bauman, J. F. (2017). Vocabulary and reading comprehension: The nexus of meaning. In: S. E. Israel (Ed.), *Handbook of Research on Reading Comprehension* (pp. 377-405). New York: The Guilford Press.

Ashrafzadeh, A., Moh Don, Z., Meshkat, M. (2015). The effect of familiarity with content knowledge on Iranian medical students' performance in reading comprehension texts: A comparative study of medical and TEFL Students. *Journal of Language Teaching and Research*, 6(3), 524-534.

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bailin, A. & Grafstein A. (2016). *Readability: Text and context*. New York: Palgrave Macmillan.

Barry, S. & Lazarte, A. A. (1998). Evidence for mental models: How do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *The Modern Language Journal* 82(2), 176-193.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bengeleil, N. F., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *The Canadian Modern Language Review*, 61(2), 225-249.

Bernhardt, E. (2011). *Understanding advanced second-language reading*. New York: Routledge.

Bernhardt, E. B. (1986). Cognitive processes in L2: An examination of reading behaviors. In J. Lantolf & A. Labarca (Eds.), *Research in second language acquisition in the classroom setting* (pp. 35-51). Norwood, New Jersey: Ablex Publishing.

Bernhardt, E. B. (1986). Reading in the foreign language. Listening, reading, and writing. *Analysis and application*, 93-115.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. Norwood, New Jersey: Ablex Publishing.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20(3), 463-494.

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York: Routledge.

Bransford, J. D. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717-726.

Brantmeier, C. (2003). Beyond linguistic knowledge: Individual differences in second language reading. *Foreign Language Annals*, 36(1), 33-43.

Bråten, I., & Strǿmsǿ, H. (2003). A longitudinal think-aloud study of spontaneous strategic processing during the reading of multiple expository texts. *Reading and Writing: An Interdisciplinary Journal*, 16, 195-218.

British Educational Association (BERA) guidelines https://www.bera.ac.uk/resources/all-publications/resources-for-researchers

Brunfaut, T. & McCray, G. (2015). *Looking into test-takers' cognitive processes while completing reading tasks: A mixed method eye-tracking and stimulated recall study*. British Council ARAGs research reports online. Online version. Retreived from https://www.britishcouncil.org/sites/default/files/brunfaut_and_mccray_report_final_0.pdf

Brunfaut, T. (2014). Language for specific purposes: Current and future issues. *Language Assessment Quarterly*, 11, 216-225.

Brunfaut, T (2022). Assessing Reading. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing.* (2nd ed.) (pp. 254-267). London: Routledge.

Bügel, K. & Buunk, B. P. (1996). Sex differences in foreign language text comprehsnion: The role of interests and prior knowledge. *The Modern Language Journal*, 80(1), 15-31.

Burgoyne, K., Whiteley, H. E. & Hutchinson, J. M. (2013). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *British Journal of Educational Psychology*, 81, 344-354.

Cain, K., Oakhill, J. V., Barnes, M. Y. & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29 (6), 850-859.

Carrell, P. L. & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17(4), 553-573.

Carrell, P. L. & Wise, T. E. (1997). The relationship between prior knowledge and topic interest in second language reading. *SSLA*, 20, 285-309.

Carrell, P. L. (1988). Some causes of text-boundedness and schema interference in ESL reading. In P. L. Carell, J. Devine & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 101-113). Cambridge: Cambridge University Press.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives*. 2nd edition. New York: Routledge.

Cavalcanti, M. C. (1987). Investigating FL reading performance through pause protocols. In C. Faerch & G. Kasper (Eds.), *Introspection in Second Language Research* (pp. 230-250). Clevedon: Multilingual Matters.

Cervetti, G. N. & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In E. Birr Moje, P. P. Afflerbach, P. Enciso & N. K. Lesaux (Eds.), *Handbook of Reading Research, Volume V*. (pp. 237-260). New York: Routledge.

Chalhoub-Deville, M. & O'Sullivan, B. (2020). *Validity. Theoretical development and integrated arguments.* British Council monographs on Modern Language Testing. Shefield: Equinox Publishing.

Chang, C. (2006). Effects of topic familiarity and linguistic difficulty on the reading strategies and mental representations of nonnative readers of Chinese. *Journal of Language and Learning*, 4(2), 172-198.

Chapelle, C. A. & Lee, H-w (2022). Conceptions of validity. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 17-31). London: Routledge.

Chen, Q. & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal* 81, 209-227.

Chigayeva, S. (2001). *The effects of discipline-specific background knowledge on reading comprehension*. Master's thesis, Iowa State University: Ames, Iowa.

Chou, M. H. (2013). Strategy use for reading English for general and specific academic purposes in testing and nontesting contexts. *Reading Research Quarterly*, 48(2), 175-197.

Chou, P. T. M. (2011). The effects of vocabulary knowledge and background knowledge on reading comprehension of Taiwanese EFL students. *Electronic Journal of Foreign Language Teaching*, 8(1), 108-115.

Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension.* Cambridge: University of Cambridge Press.

Clapham, C. (2001). Discipline specificity and EAP. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 84-100). Cambridge: Cambridge University Press.

Clarke, M. A. (1980). The short circuit hypothesis of ELS reading – or when language competence interferes with reading performance. *Modern Languages Journal*, 64, 203-209.

Coady, J. (1979). A Psycholinguistic model of the ESL Reader. In R. Mackay, B. Barkman & R. R. Jordan (Eds.), *Reading in a second language* (pp. 5-12). Rowley: Newbury House.

Cohen, A. D. (1987). Using verbal reports in research on language learning. In C. Faerch, & G. Kasper (Eds.), *Introspection in second language research* (pp. 82-95). Multilingual Matters.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Comer, J. W. (2012). Lexical inferencing in reading L2 Russian. *Reading in a Foreign Language,* 24 (2), 209-230.

Common European framework of reference for languages: Learning, teaching, assessment (2001). Cambridge: Cambridge University Press.

Cotton, D., Falvey, D. & Kent, S. (2001). *Market leader upper intermediate*, New Edition. Harlow: Pearson Education Limited.

Crain-Thoreson, C., Lippman, M. Z. & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology*, 89(4), 579-591.

Creswell, J. W. (2014). *Research design*. Los Angeles: Sage.

Cromley, J. G., & Wills, T. W. (2016). Flexible strategy use by students who learn much versus little from text: transitions within think-aloud protocols. *Journal of Research in Reading,* 39(1), 50-71.

Cronbach, L.J. (1988). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence, measurement theory and public policy* (pp. 147-171). University of Illinois Press.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

Davis, J. N. & Bistodeau, L. (1993). How do L1 and L2 reading differ? Evidence from think aloud protocols. *The Modern Language Journal*, 77(4), 459-472.

Douglas, D. & Myers, R. (2000). Assessing the communication skills of veterinary students: Whose criteria? In J. A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 60-81). Cambridge: Cambridge University Press.

Douglas, D. (2001). Language for specific purposes assessment criteria: where do they come from? *Language Testing*, 18(2). 171-185.

Douglas, D. (2001). Three problems in testing language for specific purposes: Authenticity, specificity and inseparability. In C. Elder, A. Brown, E. Grove, K. Hill, T. Iwashita, T Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty* (pp. 45-52). Cambridge: Cambridge University Press.

Droop, M. & Verhoeven, L. (1998). Background knowledge, linguistic complexity and second-language reading comprehension. *Journal of Literacy Research*, 30(2), 253-271.

Duke, N. K, Pearson, P. D., Strachan, S. L. & Billman, A. K. (2011). Essential Elements of Fostering and Teaching Reading Comprehension. In S. J. Samuels & A. E. Farstrup (Eds.), *Reading instruction* (pp. 51-93). Newark, DE: International Reading Association.

Eidswick, J. (2010). Interest and prior knowledge in second language reading comprehension. *JALT Journal*, 32(2), 149-168.

Elbro, C. & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, 17, 435-452.

Erçetin, G. (2010). Effects of topic interest and prior knowledge on text recall and annotation use in reading a hypermedia text in the L2. *ReCALL*, 22(2), 228-246.

Ericsson, A. K. and Simon, H. A. (1987). Verbal Reports on Thinking. In C. Faerch, & G. Kasper (Eds.), *Introspection in second language research* (pp. 24-53). Multilingual Matters.

Ericsson, K. A., & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Eysenck, M. W. & Keane, M. T. (1995). *Cognitive Psychology*. Hove: Psychology Press.

Fogarty, R. J., Kerns, G. M. & Pete, B. M. (2021). *Literacy reframed.* Bloomington: Solution Tree Press.

Fox, E. (2020). Readers' individual differences in affect and cognition. In E. Birr Moje, P. P. Afflerbach, P. Enciso & N. K. Lesaux (Eds.), *Handbook of reading research, volume V* (pp. 180-196). New York: Routledge.

Ghaith, G. & Obeid, H. (2004). Effect of think alouds on literal and higher–order reading comprehension. *Educational Research Quarterly*, 27(3), 49-57.

Goldman, S. R. & Duran, R. P. (1988). Answering questions from oceanography texts: Learner, task and text characteristics. *Discourse Processes*, 11(4), 373-412.

Goodman, K. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6, 126-35.

Gough, P. B. (1972). One Second of Reading. *Visible Language*, 6(4), 291-320.

Grabe, W. P. & Stoller, F. L. (2013) *Teaching and researching reading*. New York: Routledge.

Grabe, W. P. & Stoller, F. L. (2019) Reading to learn. Why and how content-based instructional frameworks facilitate the process. In K. Koda & J. Yamashita (Eds.), *Reading to learn in a foreign language* (pp. 9-30). New York: Routledge.

Grabe, W. P. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25, 375-406.

Grabe, W. P. (2000). Reading research and its implications for reading assessment. In J. A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 226-262). Cambridge: Cambridge University Press.

Grabe, W. P. (2009). *Reading in a second language. Moving from theory to practice*. Cambridge: Cambridge University Press.

Graeasser, A. C., Gordon, S. E. & Sawyer, J. D. (1979). Recognition for typical and atypical actions in scripted activities. *Journal of Verbal Learning and Verbal Behavior* 18, 319-333.

Green, R. (2000). *An empirical investigation of the componentiality of EAP reading and EAP listening through language test data.* Unpublished PhD thesis. University of Reading, Centre for Applied Language Studies.

Green, R. (2020). Pilot testing: Why and how we trial. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (115-124). New York: Routledge.

Gross, R. (1996). *Psychology.* Abingdon: Hodder & Stoughton.

Gunderson, L., D'Silva, R. A. & Murphy Odo, D. (2020). *ESL (ELL) Literacy instruction.* New York: Routledge.

Guthrie, J., Britten, T. & Barker, K. (1991). Roles of document structure, cognitive strategy and awareness in searching for information. *Reading Research Quarterly*, 26, 300-324.

Haarstrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen: Gunter Narr Verlag.

Halliday, M. A. K & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hammadou, J. A. (2000). The impact of analogy and content knowledge on reading comprehension: What helps, what hurts. *The Modern Language Journal,* 84, 38-50.

Hatch, E. & Lazaraton, A. (2000). *The research manual – design & statistics for applied linguistics.* Boston: Henile & Heinle Publishers.

Hennessy, N. L. (2021). *The reading comprehension blueprint*. Baltimore: Paul H. Brookes Publishing Co.

Hill, Y. Z., & Liu, O. L. (2012). Is there any interaction between background knowledge and language proficiency that affects TOEFL iBT® reading performance? *TOEFL iBT® Research Report.* TOEFL iBT-18, 1-34.

Hirsch, E. D. Jr. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, spring, 10-45.

Hoover, W. A. & W. E. Tunmer. (1993). The components of reading. In J. B. Thompson, W. E. Tunmer & T. Nicholson (Eds.), *Reading Acquisition Processes* (pp. 1-19). Clevedon: Multilingual Matters.

Hopkins, W.G. (2001). *A new view of statistics*. A scale of magnitudes for effect statistics http://www.sportsci.org/resource/stats/effectmag.html#cohen

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language*, 27(1), 22-46.

Hosenfeld, C. (1984). Case studies of ninth grade readers. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 231-249). London: Longman.

Hu, H. C. M. & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies and their relationship with the retention of word meanings inferred from context. *Canadian Modern Language Review*, 68(1), 54-77.

Hu, H. C. M. & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferencers. *System*, 45, 27-38.

Hudson, T. (1982). The effects of induced schemata on the 'short circuit' in L2 reading: non-decoding factors in L2 reading performance. *Language Learning*, 32, 2-31.

Hung, C. M. (1990). The effect of the pre-reading instruction on the comprehension of text by ESL readers. *The English Teacher*, 9, 1-9.

Hwang, H. J. & Duke, N. K. (2020). Content counts and motivation matters: Reading comprehension in third-grade students who are English learners. *AERA Open*, 6(1), 6-11.

Hwang, H. J. (2019). The role of science domain knowledge and reading motivation in predicting informational and narrative reading comprehension in L1 and L2: An international study. *Learning and Individual Differences*, 76, 1-11.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes,* 21, 385-395.

Hynd, C. R. & Alverman, D. E. (1989). Overcoming misconceptions in science: An on-line study of prior knowledge activation. *Reading Research and Instruction*, 28 (4), 12-26.

Israel, S. E. (2015). *Verbal protocols in literacy research.* New York: Routledge.

Israel, S. E. (Ed.) (2017). *Handbook of research on reading comprehension*. New York: The Guilford Press.

Jackson V. (2016). Applying the think-aloud strategy to improve reading comprehension of science content. *Current Issues in Education*, 19(2), 1-35.

Ja'far, W. M. (1992). *The interactive effects of background knowledge on ESP reading comprehension proficiency tests*. Unpublished PhD Thesis. University of Reading.

Jamalipour, S. & Farahani A. A. K. (2015). The effect of vocabulary knowledge and background knowledge on Iranian EFL learners’ L2 reading comprehension. *Journal of Applied Linguistics and Language Research*, 2(2), 107-121.

Jensen, C. & Hansen, C. (1995). The effect of prior knowledge on EAP listening Test performance. *Language Testing* 12(1) 99-119.

Jeon E. H. & Yamashita, J. (2020). Measuring L2 Reading. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (265-274). New York: Routledge.

Jimenez, R. T., Garcia, G. E. & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, 3(1), 90-112.

Johnson, P. (1981). Effects on reading comprehension of language complexity and cultural background of a text. *TESOL Quarterly* 15, 169-81.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: from eye fixation to comprehension. *Psychological Review,* 87, 326-354.

Just, M. A. & Carpenter, P. A. (Eds.) (1977). *Cognitive processes in comprehension*. New York: Lawrence Erlbaum.

Just, M.A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston: Allyn and Bacon.

Kaakinen, J. K. & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eye-tracking, and recall. *Discourse processes*, 40(3), 239-257.

Kaivanpanah, S. & Moghaddam, M. S. (2012). Knowledge sources in EFL learners' lexical inferencing across reading proficiency levels. *RELC Journal*, 43(3), 373-391.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-35.

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4^{th} ed.) (pp. 17-64). Westport, CT: American Council on Education ana Praeger.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73.

Kane, M. T. (2022). Articulating a validity argument. In *G. Fulcher & L. Harding (Eds.), The Routledge handbook of language testing* (2nd ed.) (pp. 32-47). London: Routledge.

Karimi, M. N. (2017). The mediated/unmediated contributions of language proficiency and prior knowledge to L2 multiple-texts comprehension: A structural equation modelling analysis. *Applied Linguistics*, 39(6), 912-932.

Karimi, M. N. (2018). Prior topical knowledge and L2 proficiency as determinants of strategic processing in English for academic purposes multi-texts comprehension. *Innovation in language learning and teaching*. 12(3) 235-246.

Kelly E. M. (2014). *The relationship between readers' prior knowledge and comprehension of expository texts*. Unpublished Master's Thesis, State University of New York at Fredonia, New York.

Kendeou, P. & O'Brien, E. J (2016). Individual differences in prior knowledge. In P. Afflerbach (Ed.), *Handbook of individual differences in reading* (pp.151-163). New York: Routledge.

Kendeou, P. & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, 97(2), 235-245.

Khalifa, H. & Weir, C. J. (2009). *Examining reading research practice in assessing second language reading.* Cambridge: Cambridge University Press.

Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363-394.

Kintsch, W. (1998). Comprehension – a paradigm for cognition. Cambridge: Cambridge University Press.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing*, 19(2), 193-220.

Koda, K. (2005). *Insights into second language reading. A cross-linguistic approach*. Cambridge: Cambridge University Press.

Koda, K. (2020). Theory-guided reading instruction in second-language classrooms. In N. Polat, T. Gregersen & P. D. MacIntyre (Eds.) *Research-driven pedagogy*. (37-54), New York: Routledge.

Koh, M. Y. (1985). The role of prior knowledge in reading comprehension. *Reading in a Foreign Language*, 3, 375-80.

Krekeler, C. (2006). Language for special academic purposes (LSAP) testing: the effect of background knowledge revisited. *Language Testing*, 23(1), 99-130.

Lahuerta Martinez, A. C. (2013). The use of various assessment tasks in the analysis of the effect of prior knowledge and interest on L2 reading comprehension. *RESLA*, 26, 289-306.

Lee, H. C. (2014). Inferencing behaviour of ESL readers. *Electronic Journal of Foreign Language Teaching*, 11(1), 21-37.

Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87-118.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: topic familiarity and working memory. *Language Learning*, 57(2), 229-270.

Leow, R. P. & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26(1), 35-57.

Li, M. & D'Angelo, N. (2016). Higher-level processes in second language reading comprehension. In X. Chen, V. Dronjic & R. Helms-Park (Eds.), *Reading in A Second language* (pp. 159-194). New York: Routledge.

Lin, Y. H. & Chern, C. L. (2014). The effects of background knowledge and L2 reading proficiency on Taiwanese university students' summarization performance. *Contemporary Educational Research Quarterly*, 22(4), 149-186.

Lin, Z. (2002). Discovering EFL learners' perception of prior knowledge and its roles in reading comprehension. *Journal of Research in Reading,* 25(2), 172-190.

Loewen, S. & Plonsky, L. (2016). *An A-Z of applied linguistics research methods*. New York: Palgrave Macmillan.

Mahmoudi, S. & Mahmoudi A. (2017). Can topic familiarity override language proficiency in reading comprehension? *Journal of Language Teaching and Research*, 8(3), 496-506.

Mandler, J. M. & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology,* 9, 111-151.

Martinez, L. C. (2009). A state-of-the-art review of background knowledge as one of the major factors that influence reading comprehension performance. *ELIA* 9, 31-57.

McCray, G. & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, 35(1), 51-73.

McGraw, K. O., & Wong, S. P. (1992). A common language effect-size statistic. *Psychological Bulletin*, 111, 361-365.

McNeil, L. (2011). Investigating the contributions of background knowledge and reading comprehension strategies to L2 reading comprehension: an exploratory study. *Reading and Writing*, 24, 883-902.

McNeil, L. (2012). Extending the compensatory model of second language reading. *System,* 40, 64-76.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement,* 3rd ed. (pp.13-103). New York: American Council on Education & Macmillan.

Meyers, J., Lytle, S. Palladino, D. Devenpeck, G. & Green, M. (1990). Think-aloud protocol analysis: An Investigation of reading comprehension strategies in fourth- and fifth-grade students. *Journal of Psychoeducational Assessment*, 8, 112-127.

Mikeska, J. N. (2010). *Towards a more complex view of genre: the importance of prior knowledge and the nature of the science content for understanding elementary* *students' comprehension of informational and data science texts*. Unpublished PhD Thesis. Michigan State University.

Mohammed, M. A. H. & Swales, J. M. (1984). Factor affecting successful reading of technical instructions. *Reading in a Foreign Language*, 2, 206-217.

Nalliveettil, G. M. (2014). Assessing reading strategies of engineering students: Think aloud approach. *English Language Teaching*, 7(5), 38-49.

Nassaji, H. (2003). L2 vocabulary learning from context: strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly,* 37, 645-70.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, 90(3), 387-401.

Nassaji, H. (2007). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, 57(1), 79-113.

Nation, I. S. P. & Macalister, J. (2021). *Teaching ESL/EFL reading and writing*. New York: Routledge.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2008). *Teaching vocabulary strategies and techniques*. Boston: Heinle.

Nuttall, C. (1996). *Teaching reading skills in a foreign language*. London: Heinemann.

O'Sullivan, B. (2011). Theories and practices in language testing. In P. Powel-Davies (Ed.), *New directions: Assessment and evaluation – A collection of papers* (15-24). London/East Asia: British Council.

O'Sullivan, B. (2012). Assessment issues in languages for specific purposes. *The Modern Language Journal*, 96, Focus Issue, 71-88.

O'Sullivan, B. (2014). *Stakeholders and consequence in test development and validation*. Plenary Address at the Language Testing Forum, University of Southampton.

O'Sullivan, B. (2016). Validity: What is it and who is it for? In Yiunam Leung (Ed.), *Epoch making in English teaching and learning: Evolution, innovation, and revolution.* Taipei: Crane Publishing Company Ltd.

O'Sullivan, B. & Weir, C. J. (2002). *Research issues in testing spoken language.* Report for the University of Cambridge Local Examination Syndicate. London: University of Surrey, Roehampton.

O'Sullivan, B. & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language Testing: Theories and Practices* (pp. 13-32). Basingstoke: Palgrave Macmillan.

Oded, B. & Stavans, A. (1994). The effect of "false" schema activation on the construction of meaning. *System*, 22(4), 497-507.

Ogle, D. M. (1986). K-W-L: A teaching model that develops active reading of expository text. *The Reading Teacher*, 39(6), 564-570.

Olshavsky, J. E. (1977). Reading as a problem solving: An investigation of strategies. *Reading Research Quarterly*, 12, 654-674.

Olson, G. M., Duffy, S.A. & Mack, R. L. (1984). Thinking-out loud as a method for studying real-time comprehension processes. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension* (pp. 253-283). New York: Laurence Erlbaum Associates.

Olson, G. M., Mack, R. L. & Duffy, S. A. (1981). Cognitive aspects of genre. *Poetics*, 10, 283-315.

Oostendorp, H. van & Goldman, S. R. K (1999). Introduction: Some initial considerations. In H. van Oostendorp & S. R. K. Goldman (Eds.), *The construction of mental representations during reading* (pp. vii-xiv). Mahwah: Laurence Erlbaum Associates.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Printer Publishers.

Paran, A. 1996. Reading in EFL: Facts and fictions, *ELT Journal* 50(1), 25-33.

Pearson, P. D. & Cervetti, G. N. (2017). The roots of reading comprehension instruction. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (pp. 12-56). New York: The Guilford Press.

Peretz, A. S. & Shoham, M. (1990). Testing reading comprehension in LSP: Does topic familiarity affect assessed difficulty and actual performance? *Reading in a Foreign Language*, 7(1), 447-455.

Perfetti, C. A., Rouet, J. F. & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostedorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). New York: Psychology Press.

Pill, J. & Smart, C. (2020). Raters: Behavior and training. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (135-144). New York: Routledge.

Plonsky, L. & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.

Pressley, M., & Afflerbach, M. (1995). *Verbal protocols of reading. The nature of constructively responsive reading.* New Jersey: Lawrence Erlbaum Associates.

Pritchard, R. & O’Hara, S. (2006). Using think alouds to identify an teach reading comprehension strategies. *The CATESOL Journal*, 18(1), 151-159.

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66-86.

Pulido, D. (2009). Vocabulary processing and acquisition through reading: Evidence for the rich getting richer. In Z. Han & N. J. Anderson (Eds.), *Second language reading research and instruction* (pp. 65-82). Michigan: University of Michigan Press.

Rayner, K., Pollatsek, A., Ashby, J. & Clifton, C. (2011). *The Psychology of Reading*. New York: Psychology Press.

Révész; A. & Brunfaut, T. (2020). Validating assessments for research purposes. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (21-32). New York: Routledge.

Ridgway, T. (1997). Thresholds of the background knowledge effect in foreign language reading. *Reading in a Foreign Language,* 10, 55-76.

Roller, C. M. & Matambo, A. R. (1992). Bilingual readers ' use of background knowledge in learning from text. *TESOL Quarterly* 26, 129-41.

Rouhi, A. & Ashgari, F. (2011). The effect of topic and pre-reading activities on reading comprehension and lexical inferencing. *Hong Kong Journal of Applied Linguistics*, 13(1), 27-40.

Ruddell, R. B., Unrau, N. J. & McCormick, S. (2019). A sociocognitive model of meaning-construction. In D. E. Alvermann, N. J. Unrau, M. Sailors & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (pp. 204-232). New York: Routledge.

Rumelhart, D. E. & Norman, D. A. (1983). Representations in memory. In Atkinson, R. C. (Ed.) *Technical report*.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic, (Ed.), *Attention and performance* (pp. 573-603). New York: Routledge.

Rydland, V., Aukrust, V. G. & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Reading and Writing*, 25, 465-482.

Sadoski, M., Paivio, A. & Goetz, E. T. (1991). A critique of schema theory in reading and a dual coding alternative. *Reading Research Quarterly*, 26 (4), 463-489.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. London: Sage Publications.

Salmani-Nodoushan, M. A. (2003). Text familiarity, reading tasks, and ESP test performance: A study on Iranian LEP and non-LEP university students. *The Reading Matrix*, 3(1), 1-14.

Sarig, G. (1987). High-level reading in the first and in the foreign language: Some comparative process data. In J. Devine, P. Carrel & D. E. Eskey (Eds.), *Research in reading in English as a second language* (pp.105-120). Washington D.C.: TESOL.

Schank, R. C. & Abelson, R. (1977). Scripts, plans and knowledge. In P. N. Johnson-Laird & P. C. Wason (Eds), *Thinking – reading in cognitive science* (pp. 421-431). Cambridge: Cambridge University Press.

Schank, R. C. (1980). What's a schema anyway? Review of R. O. Freedle (Ed.) New directions in discourse processing. *Contemporary Psychology* 25, 814-816.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary*. London: Palgrave MacMillan.

Shanahan, C. (2009). Disciplinary comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 240-260). New York: Routledge.

Shanahan, C. (2017). Comprehension in the disciplines. In: S. E. Israel (Ed.), *Handbook of research on reading comprehension* (pp. 479-499). New York: The Guilford Press.

Shaw, P & McMillion, A. (2018). Reading comprehension in advanced L2 readers. In K. Hylenstam, I. Barting & L. Fant (Eds.), *High-level language proficiency in second language and multilingual contexts* (pp. 146-169). Cambridge: Cambridge University Press.

Shin, J., Dronjic, V., & Park, B. (2019). The interplay between working memory and background knowledge in L2 reading comprehension. *TESOL Quarterly*, 53(2), 320-347.

Shiotsu, T. (2010). *Components of L2 reading*. Cambridge: Cambridge University Press.

Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read.* New York: Holt Rinehart & Winston.

Song T. & Reynolds B. L. (2022). The effects of lexical coverage and topic familiarity on the comprehension of L2 expository texts. *TESOL Quarterly*, 56(2), 763-774.

Spires, H. A. & Donley, J. (1998). Prior knowledge activation: Inducing engagement with informational texts. *Journal of Educational Psychology*. 90(2), 249-260.

Stahl, A. S., Chou Hare, V. & Sinatra, R. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension: The retiring of number 41. *Journal of Literacy Research*, 23(4), 487-508.

Stanovich, K. E. (1980). Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly* 16(1), 32-71.

Steffenson, M. S., Joag-Dev & Anderson, R. C. (1979). A cross cultural perspective on reading comprehension. *Reading Research Quarterly* 15(1), 10-29.

Stevenson, M., Schoonen, R. & de Glopper, K. (2007). Inhibition or compensation? A multidimensional comparison of reading processes in Dutch and English. *Language Learning*, 57(1), 115-154.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education* 48, 1273–1296.

Taboada, A. & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to the construction of knowledge from reading information text. *Journal of Literacy Research*, 38(1). 1-35.

Tavakoli, M. & Hayati, S. (2011). The relationship between lexical inferencing strategies and L2 proficiency of Iranian EFL Learners. *Journal of Language Teaching and Research*, 2(6), 1227-1237.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9 (1), 77-110.

Trabasso, T. & Sperry, L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language* 24 (5), 595-611.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through online inferences and mental operations in working memory. *Discourse Proceses*; 16(1-2), 3-34.

Trace, J. (2019). Clozing the gap: How far do cloze items measure? *Language Testing*, 37(2), 235-253.

Urquhart, A. H. & Weir, C. J. (1998). *Reading in a second language: Process, Product and Practice*. London: Longman.

Uso-Juan E. (2006). The compensatory nature of discipline-related knowledge and English-language proficiency in reading English for academic purposes. *The Modern Language Journal*, 90(2), 210-227.

van den Broek, P. & Kenedeou, P. (2008). Cognitive processes in comprehension or science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22, 335-351.

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

van Loon, M. H., de Bruin, van Gog, T. & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*. 24, 15-25.

Wade, S. E., Trathen, W. & Schraw, G. (1990). Analysis of spontaneous study strategies. *Reading Research Quarterly*, 25(2), 147-166.

Wallace, C. (1992). *Reading. Oxford*: Oxford University Press.

Webb, S. & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Weir, C. J. (1983). *Identifying the language needs of overseas students in tertiary education in the United Kingdom.* Unpublished PhD Thesis. University of London, Institute of Education.

Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

Weir, C. J. (2005). *Language Testing and Validation*. Basingstoke: Macmillan.

Widowson, H. G. (1979). *Explorations in Applied Linguistics.* Oxford: Oxford University Press.

Yamashita, J. (2002). Reading strategies in L1 and L2: Comparison of four groups of readers with different reading ability in L1 and L2. *International Journal of Applied Linguistics*, 135(1), 1-35.

Zwaan, R. A. & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes*, 21(3), 289-327.

Appendix A-1: TEEP Test of Grammar

Tick one item (A, B, C or D) that best fits the blank.

1. "Let's go for a coffee some time."
   "Good idea: What ..... after the lecture?"
   A) shall you do B)you doing C) you do D) are you doing

2. "We ..... to his lecture yesterday."
   A) went B) have gone C) are gone D) go

3. "It's highly inflammable, so you ..... better be careful."
   A) had B) would C)should D) must

4. "..... you need some help with the experiment, just tell me."
   A) Will B) Do C) Would D) Should

5. We wrote down the results which were ..... during the practical.
   A) observed B) happened C) occurred D)arrived

6. The lecturer informed .....
   A) his decision to us. B)us of his decision. C) his decision for us. D) us his decision.

7. "I am used ..... late at night."
   A) work B) that I work C) to working D) I work

8. The ..... industry is very important at the present time to our economy.
   A) oil B) foreign C) modern D) light

9. ..... of the students has started the course.
   A) Several B) Both C)Neither D) Most

10. The metal was ..... hot that he couldn't touch it.
    A) very B) too C) so D) extremely

11. He did not care ..... or not the man was innocent.
    A) whichever B) if C) whether D) however

12. You have to allow ..... a margin of error.
    A) by B) with C) to D) for

13. "Please turn your stereo down. I ..... to study."
    A) am trying B) have tried C) tried D) try

14. "By the time this course finishes ..... a lot about engineering."
    A) I will learn B) I learn C) I will have learnt D) I have learnt

15. "Whatever has happened to the lecturer?"
    "I don't know. He ..... lost."
    A) can have got B) could get C) might get D) may have got

16. If you have read the notes you ..... the answer.
    A) should know B) have known C) shall know D) would know

17. In this test you have to do ..... a dictionary.
    A) without B) with C) for D) by

18. This exercise ..... us with a number of interesting possibilities.
    A) sets B) shows C) gives D) presents

19. We found ..... to understand his lecture.
    A) difficulty B) difficult C) so difficult D) it difficult

20. "John's very friendly. He's from ..... England."
    A) the north B) the north of C) north the D) north of the

21. This year the number of candidates who ..... not been worked out.
A) are successful have B) are successful has C) is successful have D) is successful has

22. My research findings were not ..... to be published.
A) interesting so B) interesting enough C) enough interesting D) so interesting

23. We ..... have not discovered what causes certain illnesses.
A) yet B) already C) still D) to now

24. P.T.O. stands ..... "Please Turn Over".
A) as B) like C) for D) by

25. Coming to study in a foreign country may not be east at first, but most people eventually enjoy ..... a student in Britain.
A) being B) be C) been D) to be

26. As a result of his lectures she ..... by this new approach to teaching.
A) was influenced B) has influenced C) influenced D) had influenced

27. "I ..... yesterday, so I couldn't go to the cinema."
A) must study B) must have studied C) had to study D) ought to study

28. "If he had known the problems, he ..... the task."
A) will not have undertaken B) had not undertaken C) should not undertake D) would not have undertaken

29. The library ..... nearly two million volumes.
A) consists B) compares C) composes D) contains

30. We had to ..... our colleagues what we had discovered.
A) report B) say C) describe D) tell

31. "I'm going out unless you would prefer me ..... here."
A) to stay B) will stay C) that I stay D) stay

32. "..... a pity you did not check the figures with your partner."
A) What's B) That's C) There's D) It's

33. Everyone ..... a difficult course to follow.
A) believe that statistics is B) believe that statistics are C) believes that statistics are
D) believes that statistics is

34. It ..... produced the same result whenever we tried it.
A) ever B) often C) always D) every time

35. Hard working ..... he was, he did not meet the required standard.
A) which B) whilst C) though D) although

36. "My results are the same ..... yours."
A) that B) as C) than D) like

37. Writing this report is ..... very time consuming.
A) seeming B) having C) looking D) proving

38. Until I studied the disease, I ..... its cause.
A) do not understand B) have not understood C) did not understand D) cannot understand

39. His examination results were not as bad as they ..... been.
A) need have B) might have C) can have D) must have

40. "Unless I ..... be late."
A) run, I'll B) don't run, I won't C) don't run, I'll D) run, I won't

41. "I ..... to finish my thesis next year."
A) intend B) think C) decide D) will

42. The department was ..... to his application.
A) unaware B) opposed C) contrary D) uncaring

43. "You'd better ..... to the doctor next time you feel ill."
A) to go B) going C) go D) gone

44. "..... I need is a long holiday."
A) What B) That C) Which D) The which

45. A number of the areas the professor ..... in the field of atomic physics.
A) have specialised in are B) has specialised in is C) have specialised in is
D) has specialised in are

46. He is ..... proud man that he would rather fail than ask for help.
A) so a B) such C) a so D) such a

47. ..... Newton, "to every action there is an equal and opposite reaction".
A) Because of B) According to C) By D) In contrast

48. "I am taller than you ..... three inches."
A) with B) by C) of D) in

49. "Have you finished your project?" "Yes, it ..... now."
A) typed B) is being typed C) types D) typing

50. "Your English is very good." "It should be. I ..... it ever since I started school."
A) have been learning B) was learning C) had learned D) had been learning

51. "I wonder why he didn't come to class." "He ..... his bus."
A) can have missed B) could miss C) may have missed D) might miss

52. If only he ..... down the results when he did the experiments!
A) writes B) had written C) has written D) was writing

53. He ..... them do the experiment again.
A) insisted B) allowed C) made D) requested

54. The university was ..... to those who had difficulty paying their fees.
A) pitiful B) hopeful C) sympathetic D) punishable

55. Caramel is a brown substance ..... by the action of heat on sugar.
A) form B) forming C) formed D) forms

56. "Would you like some more meat? There's still ..... left."
A) a little B) little C) a few D) few

57. Not all the students ..... criteria for assessing written work.
A) understands these B) understands this C) understand these D) understand this

58. ..... how hard he worked, his tutor never commented on it.
A) Of no account B) No matter C) Without Regards D) Mindless

59. The penguin is a bird adapted to life ..... on land and in water.
A) both B) not only C) and D) either

60. ..... many years he studied hard for his doctorate.
A) During B) For C) Since D) From

Appendix A-2: Language experts' assessment of Grammar test

TEST ITEM 1

How does this item measure knowledge of grammar? Please circle.

5 Very well	4 Well	3 Neither well nor badly	2 Badly	1 Very badly

Item	Very well	Well	Neither well nor badly	Badly	Very badly	No. of raters
1	4	2	0	1	0	7
2	3	3	1	0	0	7
3	4	1	1	1	0	7
4	4	3	0	0	0	7
5	3	3	0	0	1	7
6	2	3	0	1	0	6
7	3	2	0	2	0	7
8	0	0	1	1	5	7
9	4	2	0	1	0	7
10	3	2	0	1	0	6
11	2	2	2	1	0	7
12	3	0	2	1	1	7
13	3	3	1	0	0	7
14	5	1	1	0	0	7
15	3	3	0	0	0	6
16	3	2	1	0	0	6
17	2	0	2	1	1	6
18	1	0	1	3	1	6
19	2	2	1	0	1	6
20	2	1	2	0	1	6
21	3	2	1	0	0	6
22	3	2	1	0	0	6
23	2	2	0	1	1	6
24	1	1	2	0	1	5
25	3	2	0	0	0	5
26	3	3	0	0	0	6
27	3	2	0	1	0	6
28	3	2	0	1	0	6
29	1	2	1	1	2	7
30	3	1	1	2	0	7
31	2	4	0	1	0	7
32	2	1	1	2	0	6
33	6	0	0	1	0	7
34	1	2	2	0	2	7
35	2	4	0	1	0	7
36	2	4	1	0	0	7
37	2	1	2	0	0	5
38	3	2	0	1	0	6
39	2	3	2	0	0	7
40	6	1	0	0	0	7
41	2	1	3	0	1	7
42	0	1	2	1	2	6
43	3	4	0	0	0	7

44	1	2	2	1	1	7
45	5	1	0	0	0	6
46	3	2	2	0	0	7
47	1	1	3	1	1	7
48	1	0	2	2	1	6
49	4	1	1	1	0	7
50	5	2	0	0	0	7
51	5	2	0	0	0	7
52	4	2	1	0	0	7
53	4	1	2	0	0	7
54	2	0	1	2	2	7
55	2	4	1	0	0	7
56	5	2	0	0	0	7
57	5	0	1	1	0	7
58	2	1	3	0	1	7
59	2	2	1	1	1	7
60	2	2	2	1	0	7

Items measuring grammar "very badly": 5, 8, 12, 17, 18, 19, 20, 23, 24, 29, 41, 42, 44, 47, 48, 54, 58, 59.

Items measuring grammar "badly": 1, 6, 7, 8, 9, 10, 11, 12, 17, 18, 23, 27, 28, 29, 30, 31, 32, 33, 35, 38, 42, 44, 47, 48, 49, 54, 57, 59, 60

Appendix A-3: Language experts' assessment of Grammar test (cultural and historical bias)

Is this item culturally biased? Please circle.	**YES**	**NO**	Could you comment.
Is this item historically biased? Please circle.	**YES**	**NO**	Could you comment.

ITEM	**CULTURAL BIAS**		**HISTORICAL BIAS**	
	YES	**NO**	**YES**	**NO**
1	1	5	0	6
2	0	6	0	6
3	0	6	0	6
4	0	6	0	5
5	0	6	0	6
6	0	6	0	6
7	0	6	0	6
8	2	2	1	3
9	0	5	0	5
10	0	5	0	5
11	1	4	1	4
12	0	5	0	5
13	1	4	0	5
14	0	5	0	4
15	1	2	0	3
16	0	2	0	2
17	0	3	0	3
18	0	3	0	3
19	0	3	0	3
20	2	1	1	2
21	0	3	0	3
22	0	3	0	3
23	0	3	0	3
24	0	3	0	3
25	1	2	0	2
26	0	3	0	3
27	0	3	0	3
28	0	3	0	3
29	0	4	0	4
30	0	4	0	4
31	0	4	0	4
32	0	3	0	3
33	0	4	0	4
34	0	4	0	4
35	0	3	0	3
36	0	3	0	3
37	0	3	0	3
38	0	3	0	3
39	0	3	0	3
40	0	3	0	3
41	0	3	0	3
42	1	2	1	2

43	0	4	0	4
44	0	3	0	3
45	0	4	0	4
46	0	4	0	4
47	1	3	0	4
48	1	2	0	3
49	0	3	0	3
50	1	1	0	2
51	0	2	0	2
52	0	2	0	2
53	0	2	0	2
54	0	1	0	1
55	0	2	0	2
56	0	2	0	2
57	0	4	0	4
58	0	4	0	4
59	0	4	0	4
60	0	4	0	4

Items with cultural bias: 8, 20, 25, 42, 47, 48

Items with historical bias: 8, 20, 42

Appendix A-4: Language experts' comments on construct validity, cultural and historical bias – Grammar test

ITEM	COMMENTS
1	• Obviously wrong answers. • Short dialogues are useful for testing grammar.
2	• Students wouldn't opt for b.
3	• Students are used to "you'd better" without realizing what it stands for. • Students are taught only "you'd better".
4	-
5	• This is a lexical rather than a grammatical item: all four verbs are in the same form!
6	-
7	• Why not offer also "to work". Students don't make mistakes with bare infinitive. • Another wrong alternative should be given: "to work" instead of "I work".
8	• Two answers possible. Is this testing grammar? • I'm not sure of the answer. • This is a lexical item. • This item doesn't measure grammar. • Checking lexical knowledge. • Not grammar but vocabulary. • What are we checking here: lexical phrase, use of the article?
9	• Good choice between b and c. • Students will probably understand this one as checking vocabulary.
10	• Why not changing "metal" to something else?
11	• Good choice between b and c. • This is a lexical phrase rather than a grammatical construction. • Why "he"? Is it necessary to stress "he"? • This sentence is strange, the second part should be changed.
12	• Students are used to "allow to". • Is this item testing lexical or grammatical knowledge? • This is a grammatical collocation.
13	• Good choice between a and d.
14	• Good choice between a and c. • The test probably doesn't allow a, but in spoken English c would sound too formal. • Correct answer sounds unnatural. • Clear and straightforward measurement of grammar.
15	• Good choice between a and d. • It may be inappropriate not to believe in orientation abilities of a senior lecturer. • No point using a, it should be replaced.
16	• Good choice between a, c and d.
17	• Good choice between a and b. • Lexical phrase "you have to do without".
18	• Not a baffling problem. I don't think anybody would decide for any other item but d.
19	• Adding "difficulty" opens another problem area. • Lexical phrase "to find it difficult to".

20	• Good choice between b and d. • Wouldn't John from the South of England be kind? Geographic and historical bias. • Why not Hans from Germany?
21	• I'm not sure. • An economical item; agreement well tested.
22	• Good choice between b and c. • It would make sense to offer a wrong alternative "interested enough".
23	• Why not have "till now" in d.
24	• Grammatical collocation. • Grammatical or lexical knowledge? • Grammatical collocation.
25	• In Britain? • Britain – enjoyable? • Good choice between a and d.
26	• A seems the only choice. • "He" again.
27	• I see c as a natural choice. • I would add "was studying" instead of b.
28	• Why not have "I would not undertake" instead of c? • "He" again. • The item tests if clauses well.
29	• Not a purely grammatical task. • Grammatical or lexical knowledge. • Lexis. • Testing meaning and grammar.
30	• Good choice between b and d. • Grammatical or lexical knowledge? • I find this to be a combination with testing lexis. If we wanted to test whether this is a transitive or intransitive verb, we would need to do this with one verb only.
31	• I would have "staying" instead of d.
32	• Good choice between a and d. • "It's a pity" is usually taught as a lexical phrase. • Common phrase. • This is not checking much.
33	• Puzzling for most students. • Two items. • One of the best items because it kills to birds at a stone. Economical.
34	• What's the difference between a and b? • Lexical or grammatical knowledge?
35	• Too difficult for most students. • "He" again.
36	• Good choice between b and d. • "The same as" can be taught as a lexical phrase.
37	• Unnatural sentence. • Good choice between c and d. • Lexical or grammatical knowledge? • Should give better alternatives.
38	• I would use Past Perfect. • Good choice between c and b.

	• What about Past Perfect?
39	• "He" again. • Models are problematic. They belong to grammar, but their semantic features are even more important.
40	• Bewildering to most students.
41	• I'm not sure about anybody choosing d. I would have "must" instead of something else. • Grammatical collocation.
42	• This is not grammar. • "He" again. • Grammatical collocation.
43	• -
44	• Seems a logical choice. • "What I need" is taught as a lexical phrase.
45	• Perplexing to most students.
46	• I would have had "so" without "a" for a. • "He". • Lexical phrase rather than a grammatical collocation.
47	• Is this grammar? • Lexical or grammatical knowledge? • A British scientist.
48	• Better use "for". • Grammatical collocation. • Better to include "for.
49	• Too obvious. • I would have "is typing" for d. • One more alterative should be suggested (is typed). The wrong ones are all active.
50	• Why not have "have learnt" instead of d? • Why not French?
51	• Always a problem. • "He" again.
52	• Why not have "wrote" instead of a? It looks wrong at first sight. • "He" again.
53	• Good choice between b and c. • "He" again. • Grammatical or lexical knowledge.
54	• This is a vocabulary item. • What does this mean in various school systems: words or actions? • Grammar or lexical knowledge? • Better alternatives should be given.
55	• Good choice between b and c.
56	• Always a problem. • What about vegetarians? • Clear and straightforward.
57	• Grammatical or lexical knowledge? • Grammar + vocabulary.
58	• "He" again. • Grammatical or lexical knowledge?
59	• -
60	• "He" again.

	Overall comments: • Grammar and lexis are often so intertwined that it is difficult to decide if the item focuses only on grammar. I think we should consider how the learner understands the situation. For instance, when learning grammatical collocations, we must learn the right preposition, so this then belongs to lexis. I don't find any item to be culturally or historically biased. • All sentences have male subjects. Is there any reason for this? I think all these sentences should be looked at.

Appendix A-5: Summary of lexicographer's assessment of Grammar test items

- **Lexical items:** 5, 8, 19, 20, 29, 30, 32, 37, 41, 47, 53, 54, 58
- **Borderline grammar-lexical items**, including grammatical collocations: 11 12, 17, 18, 23, 24, 35, 42, 44, 46, 48 59, 60

Tick one item (A, B, C or D) that best fits the blank.

1. "Let's go for a coffee some time."
"Good idea: What ..... after the lecture?"
A) shall you do B)you doing C) you do D) are you doing

2. "We ..... to his lecture yesterday."
A) went B) have gone C) are gone D) go

3. "It's highly inflammable, so you ..... better be careful."
A) had B) would C)should D) must

4. "..... you need some help with the experiment, just tell me."
A) Will B) Do C) Would D) Should

5. **We wrote down the results which were ..... during the practical.**
A) observed B) happened C) occurred D)arrived

6. The lecturer informed .....
A) his decision to us. B)us of his decision. C) his decision for us. D) us his decision.

7. "I am used ..... late at night."
A) work B) that I work C) to working D) I work

8. **The ..... industry is very important at the present time to our economy.**
A) oil B) foreign C) modern D) light

9. ..... of the students has started the course.
A) Several B) Both C)Neither D) Most

10. The metal was ..... hot that he couldn't touch it.
A) very B) too C) so D) extremely

11. **He did not care ..... or not the man was innocent.**
A) whichever B) if C) whether D) however

12. **You have to allow ..... a margin of error.**
A) by B) with C) to D) for

13. "Please turn your stereo down. I ..... to study."
A) am trying B) have tried C) tried D) try

14. "By the time this course finishes ..... a lot about engineering."
A) I will learn B) I learn C) I will have learnt D) I have learnt

15. "Whatever has happened to the lecturer?"
"I don't know. He ..... lost."
A) can have got B) could get C) might get D) may have got

16. If you have read the notes you ..... the answer.
A) should know B) have known C) shall know D) would know

17. **In this test you have to do ..... a dictionary.**
A) without B) with C) for D) by

18. This exercise ..... us with a number of interesting possibilities.
A) sets B) shows C) gives D) presents

19. We found ..... to understand his lecture.
A) difficulty B) difficult C) so difficult D) it difficult

20. "John's very friendly. He's from ..... England."
A) the north B) the north of C) north the D) north of the

21. This year the number of candidates who ..... not been worked out.
A) are successful have B) are successful has C) is successful have D) is successful has

22. My research findings were not ..... to be published.
A) interesting so B) interesting enough C) enough interesting D) so interesting

23. We ..... have not discovered what causes certain illnesses.
A) yet B) already C) still D) to now

24. P.T.O. stands ..... "Please Turn Over".
A) as B) like C) for D) by

25. Coming to study in a foreign country may not be east at first, but most people eventually enjoy ..... a student in Britain.
A) being B) be C) been D) to be

26. As a result of his lectures she ..... by this new approach to teaching.
A) was influenced B) has influenced C) influenced D) had influenced

27. "I ..... yesterday, so I couldn't go to the cinema."
A) must study B) must have studied C) had to study D) ought to study

28. "If he had known the problems, he ..... the task."
A) will not have undertaken B) had not undertaken C) should not undertake D) would not have undertaken

29. The library ..... nearly two million volumes.
A) consists B) compares C) composes D) contains

30. We had to ..... our colleagues what we had discovered.
A) report B) say C) describe D) tell

31. "I'm going out unless you would prefer me ..... here."
A) to stay B) will stay C) that I stay D) stay

32. "..... a pity you did not check the figures with your partner."
A) What's B) That's C) There's D) It's

33. Everyone ..... a difficult course to follow.
A) believe that statistics is B) believe that statistics are C) believes that statistics are
D) believes that statistics is

34. It ..... produced the same result whenever we tried it.
A) ever B) often C) always D) every time

35. Hard working ..... he was, he did not meet the required standard.
A) which B) whilst C) though D) although

36. "My results are the same ..... yours."
A) that B) as C) than D) like

37. Writing this report is ..... very time consuming.
A) seeming B) having C) looking D) proving

38. Until I studied the disease, I ..... its cause.
A) do not understand B) have not understood C) did not understand D) cannot understand

39. His examination results were not as bad as they ..... been.
A) need have B) might have C) can have D) must have

40. "Unless I ..... be late."
A) run, I'll B) don't run, I won't C) don't run, I'll D) run, I won't

41. "I ..... to finish my thesis next year."
A) intend B) think C) decide D) will

42. The department was ..... to his application.
A) unaware B) opposed C) contrary D) uncaring

43. "You'd better ..... to the doctor next time you feel ill."
A) to go B) going C) go D) gone

44. "..... I need is a long holiday."
A) What B) That C) Which D) The which

45. A number of the areas the professor ..... in the field of atomic physics.
A) have specialised in are B) has specialised in is C) have specialised in is
D) has specialised in are

46. He is ..... proud man that he would rather fail than ask for help.
A) so a B) such C) a so D) such a

47. ..... Newton, "to every action there is an equal and opposite reaction".
A) Because of B) According to C) By D) In contrast

48. "I am taller than you ..... three inches."
A) with B) by C) of D) in

49. "Have you finished your project?" "Yes, it ..... now."
A) typed B) is being typed C) types D) typing

50. "Your English is very good." "It should be. I ..... it ever since I started school."
A) have been learning B) was learning C) had learned D) had been learning

51. "I wonder why he didn't come to class." "He ..... his bus."
A) can have missed B) could miss C) may have missed D) might miss

52. If only he ..... down the results when he did the experiments!
A) writes B) had written C) has written D) was writing

53. He ..... them do the experiment again.
A) insisted B) allowed C) made D) requested

54. The university was ..... to those who had difficulty paying their fees.
A) pitiful B) hopeful C) sympathetic D) punishable

55. Caramel is a brown substance ..... by the action of heat on sugar.
A) form B) forming C) formed D) forms

56. "Would you like some more meat? There's still ..... left."
A) a little B) little C) a few D) few

57. Not all the students ..... criteria for assessing written work.
A) understands these B) understands this C) understand these D) understand this

58. ..... how hard he worked, his tutor never commented on it.
A) Of no account B) No matter C) Without Regards D) Mindless

59. The penguin is a bird adapted to life ..... on land and in water.
A) both B) not only C) and D) either

60. ..... many years he studied hard for his doctorate.
A) During B) For C) Since D) From

Appendix A-6: Results: Grammar test

	question	key	A	B	C	D	omit	share of correct answers for each question	Corrected item-total correlation	Cronbach's Alpha if item deleted
1	5	A	75%	10%	15%			75%	0.424	0.891
1	8	A	90%	5%	5%			90%	-0.028	0.895
1	19	D	5%	10%		85%		85%	0.431	0.891
1	20	B	5%	60%		35%		60%	0.507	0.889
1	29	D	15%		5%	80%		80%	0.437	0.891
1	30	D	35%		5%	60%		60%	0.620	0.888
1	32	D		30%	5%	65%		65%	0.729	0.886
1	37	D	20%	5%	35%	40%		40%	0.371	0.892
1	47	B	5%	90%			5%	90%	0.586	0.890
1	53	C		35%	55%	5%	5%	55%	0.563	0.889
1	54	C	10%	5%	70%	10%	5%	70%	-0.044	0.897
1	58	B		95%			5%	95%	0.000	0.894
2	12	D	5%	30%	15%	50%		50%	0.632	0.887
2	17	A	100%					100%	0.000	0.894
2	18	D	40%	15%	15%	30%		30%	0.234	0.893
2	24	C	5%		90%	5%		90%	0.600	0.889
2	41	A	95%		5%			95%	0.586	0.890
2	42	B	10%	60%	15%	10%	5%	60%	0.583	0.888
2	44	A	95%				5%	95%	0.000	0.894
2	46	D		5%		90%	5%	90%	-0.209	0.896
2	48	B		80%	5%	10%	5%	80%	0.468	0.890
	1	D				100%		100%	0.000	0.894
	2	A	100%					100%	0.000	0.894
	10	C			100%			100%	0.000	0.894
	13	A	100%					100%	0.000	0.894
1	25	A	95%			5%		95%	0.111	0.894
2	26	A	95%	5%				95%	0.586	0.890
3	31	A	95%			5%		95%	0.317	0.892
4	34	C			95%	5%		95%	0.081	0.894
5	39	B		95%	5%			95%	0.228	0.893
6	52	B		95%			5%	95%	0.000	0.894
7	60	B		95%			5%	95%	0.000	0.894
8	11	C		10%	90%			90%	0.164	0.894
9	22	B		90%	10%			90%	0.667	0.889
10	23	C	5%		90%	5%		90%	0.015	0.895
11	27	C		10%	90%			90%	0.121	0.894

	Cronbach's Alpha -all 60 items	0.894
deleted item circle 1	Cronbach's Alpha	0.855
deleted item circle 1&2	Cronbach's Alpha	0.815
deleted item circle 1&2+100%	Cronbach's Alpha	0.814
deleted item circle 1&2+95%	Cronbach's Alpha	0.802
deleted item circle 1&2+90%	Cronbach's Alpha	0.779

	question	key	A	B	C	D	omit	share of correct answers for each question	Corrected item-total correlation	Cronbach's Alpha if item deleted
12	36	B		90%	5%	5%		90%	0.534	0.890
13	49	B		90%		5%	5%	90%	0.111	0.894
14	3	A	5%	5%	85%	5%		5%	0.362	0.892
15	4	D	10%	25%	15%	50%		50%	0.534	0.889
16	6	B	20%	70%		10%		70%	0.509	0.889
17	7	C	5%	10%	85%			85%	0.138	0.894
18	9	C	5%	5%	45%	45%		45%	0.576	0.888
19	14	C	40%		55%	5%		55%	0.163	0.895
20	15	D		10%	15%	75%		75%	0.564	0.889
21	16	A	80%	5%		15%		80%	0.029	0.896
22	21	B	50%	45%	5%			45%	0.261	0.893
23	28	D		10%	5%	85%		85%	0.637	0.888
24	33	D	30%	5%	10%	55%		55%	0.197	0.894
25	35	C	65%	5%	20%	10%		20%	0.224	0.893
26	38	C		35%	55%	10%		55%	0.318	0.892
27	40	A	85%			15%		85%	0.394	0.891
28	43	C	5%	10%	80%		5%	80%	0.562	0.889
29	45	D		35%		60%	5%	60%	0.285	0.893
30	50	A	85%			10%	5%	85%	0.121	0.894
31	51	C	5%	5%	75%	10%	5%	75%	0.691	0.887
32	55	C	10%	10%	75%		5%	75%	0.640	0.888
33	56	A	40%	20%	20%	15%	5%	40%	0.046	0.896
34	57	C			65%	30%	5%	65%	0.131	0.895
35	59	A	70%	10%		15%	5%	70%	0.627	0.888

deleted item	Cronbach's Alpha	0.801
circle 1 & 2 +100% +question 11, 23, 35, 59 in 60		

Appendix A-7: The list of texts for selection

1. CHAMELEON BONDS – The Economist
2. FLOATING IN THE AIR – The Economist
3. FLOTATION FEVER – The Economist
4. **STAR-STRUCK – The Economist**
5. **BONDS THAT ROCK & R – the Economist**
6. GREENMAIL – Rand Journal of Economics
7. THE TIMING OF IPO'S Journal of Financial Economics
8. FINANCIAL DECISIONS – Journal of Financial Economics
9. **TAKING STOCK – The Journal of Finance**
10. **OPEN MARKET STOCK R. – Financial Management**
11. AGENCY COSTS – Journal of Financial Economics

Appendix A-8: Text selection: Assessment instrument

Please mark your answers in the table.

STAR-STRUCK									
1. Topic specificity – To what extent is the topic of this passage specific?	Not at all	1	2	3	4	5	6	7	A very great deal
2. Text appropriacy – To what extent is the passage appropriate for undergraduate students of Corporate Finance?	Not at all	1	2	3	4	5	6	7	A very great deal
3. Topic familiarity – To what extent would your students be familiar with this particular topic?	Not at all	1	2	3	4	5	6	7	A very great deal
4. New knowledge – To what extent does the passage contain knowledge that you would not expect your students to have?	Not at all	1	2	3	4	5	6	7	A very great deal
5. Explanation of new knowledge – If there is new knowledge in the passage, to what extent is it explained (through definition, example, etc.)?	Not at all	1	2	3	4	5	6	7	A very great deal

Appendix A-9: Text selection: Expert assessment of eleven texts

1. Dr. M. Č.	Specificity	Appropriacy	Familiarity	New knowl.	Explanation
CHAMELEON BONDS	5	6	6	4	3
FLOATING IN THE AIR	5	6	4	5	6
FLOTATION FEVER	5	7	7	3	3
STAR-STRUCK	7	6	2	6	6
BONDS	7	3	5	7	7
GREENMAIL	5	3	4	5	2
THE TIMING OF IPO'S	5	4	6	5	3
FINANCIAL DECISIONS	5	3	5	4	3
TAKING STOCK	7	3	2	6	3
OP. MARKET STOCK R.	7	2	3	6	2
AGENCY COSTS	5	4	3	6	2

2. Dr. M. K.	Specificity	Appropriacy	Familiarity	New knowl.	Explanation
CHAMELEON BONDS	5	6	4	4	2
FLOATING IN THE AIR	3	6	5	4	5
FLOTATION FEVER	5	5	3	5	4
STAR-STRUCK	6	6	4	5	4
BONDS	6	6	4	5	4
GREENMAIL	6	6	6	3	7
THE TIMING OF IPO'S	6	6	5	5	6
FINANCIAL DECISIONS	6	6	6	6	6
TAKING STOCK	6	5	4	5	5
OP. MARKET STOCK R.	6	6	5	6	6
AGENCY COSTS	6	7	6	4	5

3. Dr. A. B.	Specificity	Appropriacy	Familiarity	New knowl.	Explanation
CHAMELEON BONDS	2	7	7	2	6
FLOATING IN THE AIR	3	6	4	5	5
FLOTATION FEVER	4	5	6	2	6
STAR-STRUCK	6	4	3	6	4
BONDS	5	4	3	5	5
GREENMAIL	3	6	5	2	7
THE TIMING OF IPO'S	3	5	4	5	6
FINANCIAL DECISIONS	5	5	5	5	4
TAKING STOCK	5	6	6	2	6
OP. MARKET STOCK R.	7	2	2	6	3
AGENCY COSTS	3	7	7	3	5

Appendix A-10: Text selection: Expert assessment of eleven texts – combined

EXPERT ASSESSMENT (3 raters)	**Specificity (1-7)**	**Appropriacy (1-7)**	**Familiarity (1-7)**	**New knowl. (1-7)**	**Explanation (1-7)**
CHAMELEON BONDS	5, 5, 2	6, 6, 7	6, 4, 7	4, 4, 2	3, 2, 6
FLOATING IN THE AIR	5, 3, 3	6, 6, 6	4, 5 4	5, 4, 5	6, 5, 5
FLOTATION FEVER	5, 5, 4	7, 5, 5	7, 3, 6	3, 5, 2	3, 4, 6
STAR-STRUCK	7, 6,6	6, 6, 4	2, 4, 3	6, 5, 6	6, 4, 4
BONDS	7, 6, 5	3, 6, 4	5, 4, 3	7, 5, 5	7, 4, 5
GREENMAIL	5, 6, 3	3, 6, 6	4, 6, 5	5, 3, 2	2, 7, 7
THE TIMING OF IPO’S	5, 6, 3	4, 6, 5	6, 5, 4	5, 5, 5	3, 6, 6
FINANCIAL DECISIONS	5, 6, 5	3, 6, 5	5, 6, 5	4, 6, 5	3, 6, 4
TAKING STOCK	7, 6, 5	3, 5, 6	2, 4, 6	6, 5, 2	3, 5, 6
OP. MARKET STOCK R.	7, 6, 7	2, 6, 2	3, 5, 2	6, 6, 6	2, 6, 3
AGENCY COSTS	5, 6, 3	4, 7, 7	3, 6, 7	6, 4, 3	2, 5, 5

Appendix A-11: Text selection: Expert assessment of eleven texts – means

EXPERT ASSESSMENT	**Specificity Mean**	**Appropriacy Mean**	**Familiarity Mean**	**New knowl. Mean**	**Explanation Mean**
CHAMELEON BONDS	4.00	6.33	5.66	3.33	3.66
FLOATING IN THE AIR	4.33	6.00	4.33	4.66	5.33
FLOTATION FEVER	4.66	5.66	4.66	3.33	4.33
STAR-STRUCK	**6.33**	**5.33**	**3.00**	**5.66**	**4.66**
BONDS	**6.00**	**4.33**	**4.00**	**5.66**	**5.33**
GREENMAIL	4.66	5.00	5.00	3.33	5.33
THE TIMING OF IPO’S	4.66	5.00	5.00	5.00	5.00
FINANCIAL DECISIONS	5.33	4.66	5.33	5.00	4.33
TAKING STOCK	**6.00**	**4.66**	**4.00**	**4.33**	**4.66**
OP. MARKET STOCK R.	**6.66**	**3.33**	**3.33**	**6.00**	**4.00**
AGENCY COSTS	4.66	6.00	5.33	4.33	4.00

Appendix A-12: Reading texts 1-4

Star-struck

Share in your idol's success

NEXT month Japanese investors will be able to buy into the careers of five would-be stars. The five – all female and attractive – were chosen from thousands by a firm named Japan Digital Content. It is launching a "talent fund" in which it will sell shares: ultimately, returns will depend on sales of DVDs, calendars and other paraphernalia.

The idea is not entirely new. In 1997 bankers issued bonds backed by future sales of music by David Bowie, a British singer. In exchange for the temporary music rights to 25 of his albums, Mr Bowie received $55m. Investors got an 8 % yield on bonds rated at investment grade. Offerings by other top acts soon followed. David Pullman, the financier who pioneered Bowie bonds, has since issued securities backed by royalties from John Steinbeck's literary estate and from cartoon libraries.

However, equity financing of untested projects is a novelty. The Japanese pop hopefuls are not alone. In mid-November, "Billy Dead", a film to star and be produced by Ethan Hawke, an American actor, became the first movie to issue shares, which will be tradable. Civilian Capital, an online investment bank working with Mr Hawke, is on track to close the $7.9m issue in the next three months, says its president, Barry Poltermann. Filming is due to start in 2004. Investors will have first call on all profits until their money is returned, and then a percentage of earnings – from box-office sales, sequels and merchandise.

DigiCirc, another Californian firm, is also raising money from public investors, this time for music. Investors can buy or lease the rights to songs or albums posted on DigiCirc's website by unknown musicians in return for a piece of future profits.

Is a glamorous new asset-class in the making? Not yet. Moody's, a rating agency, is considering downgrading Bowie bonds. Sales of Mr Bowie's CDs have been soggy, partly because of online piracy; and EMI, the music company that guaranteed the bonds, was itself downgraded by Moody's in March. Indeed, such securitisations have not become as fashionable as many expected. Few musicians own all of the rights to their music; fewer still convince investors that their popularity will last.

Equity investments in unknown talent are even riskier. Sean Kane of Hall Dickler, a New York law firm, thinks that issuing securities backed by pools of intellectual property – from actors, authors and musicians – might solve this problem. Until then, fans-cum-investors may boost their idols' careers and incomes; but they might not be making the cleverest of investments.

Bonds that rock and roll

From cinema tickets to parking fines, almost everything is being securitised. Does it all make sense?

Fancy investing in a security whose payoff depends on how much beer is sold in British pubs? How about a bond to be paid by collections of overdue parking fines in New York City? If you'd prefer, you can purchase the rights to a slice of the revenues from old Italian films, or the amounts raised by selling executive suites in Denver's new stadium, or the royalties earned by pop stars such as David Bowie and Rod Stewart. There is barely a cash flow anywhere, it seems, that cannot be reassembled into a bond-like security that the most conservative of investors might buy.

Putting such strange instruments together has become one of the hottest businesses on Wall Street. New issues of asset-backed securities rose by a quarter last year, to $484 billion. This year's growth rate is expected to be at least as rapid, and the value of new asset-backed securities may for the first time exceed that of traditional corporate bonds. Mostly, that is good news, as securitisation can reduce the cost of borrowing. Yet amid the enthusiasm, the risk may be overlooked. The fact that new sorts of security can be created does not always mean they should be.

It is easy to see why securitisation has become so popular. Issuers gain instant access to money for which they would otherwise have to wait months or years, and they can shed some of the risk that their expected revenues will not materialise. Investment bankers are able to finance their customers without extending loans that tie up large amounts of capital. And investors can hold a new sort of asset, less risky than unsecured bonds, so they can diversify their portfolios and thus reduce their risk.

Those, at least, are the theoretical benefits. Turning them into reality can be tricky. In fact, for all the current excitement about rock-star bonds and the like, it is striking that the vast majority of asset-backed securities are American and involve a few types of assets, such as mortgage and credit-card loans, which generate relatively predictable amounts of cash at predictable times. Much of this year's growth is expected to come from banks securitising loans that are already on their books and from an increase in American-style securitisation in other markets, especially European. Yet even that is by no means certain: two previous attempts to introduce Europeans to asset-backed securities failed, even though American investment banks spent a fortune trying to make them work.

Even traditional sorts of asset-backed securities sometimes spring nasty surprises. Investors in mortgage-backed securities have repeatedly taken hits when borrowers repaid in larger numbers than computer models had predicted. And in the past year, AutoBond Acceptance, a car lender, saw profits drop and GreenTree Financial, which lends on mobile homes, reported a loss after default and pre-payment rates were higher than expected. The securities continued to perform, because investors had purchased only a portion of the expected payments rather than the whole lot.

These failures illustrate how difficult securitisation can be even with relatively straightforward assets. It is harder still to make securities out of such things as music royalties and patent licence fees.

Most traditional securitisations involve cash flows that have been earned, but are not yet received; the main risk is that the cash will flow earlier or later than promised, or not at all.

When lots of similar receivables are bundled together in an asset-backed security, the average rates of default, late payment and pre-payment are predictable, so both issuers and investors can be relatively confident of their ability to value the security properly.

Many of the more exotic asset-backed securities, by contrast, are based not on earned-but-uncollected cash, such as monthly mortgage payments, but on forecasts of future earnings. Henry Morriello, a partner at Kaye Scholer, a law firm in New York, reckons that at least five years of sales data are necessary to have much confidence about future streams of music revenues, which is why established rockers such as Messrs Bowie and Stewart have been able to do deals; it would be much harder to predict royalties for short-lived acts, such as the Spice Girls. Even so, music fans and cinema-goers are fickle, so the past may not be a reliable guide to the future.

Some of the new financial instruments also confront investors with a problem largely absent from traditional asset-backed securities: moral hazard. In this case, that means that the issuer's own actions can have a considerable impact on the value of the securities. Once a pool of mortgages is securitised there is little that the original lender can do to alter its value. The value of PolyGram's recent $650m securitisation of ticket, video and merchandising revenues, however, depends heavily upon how hard the company promotes its upcoming films. There may be moral issues of different sort as well. "What happens to record sales if the star is caught molesting kids?" asks Paul Taylor, of Duff & Phelps, a bond-rating agency.

Odd couples

One reason investors love securities which pool assets is that they limit risk by minimising the loss if something awful happens to any one asset. This is harder to do with royalties, film revenues and patent licences: One rock star's royalty stream differs from another's in a way that two 25-year mortgages rarely do. Although Rod Stewart is reported to have securitised his royalties, that is only half true; he merely received a $15m loan secured against future royalties. Bankers hope to bundle this with other loans to entertainers and fully securitize it, but this is proving hard to do. Just negotiating the simpler deal announced last week took more than a year.

Complicating matters further, it is not always clear whether the issuer of such one-off securities actually owns the cash flow being securitised. There can be huge legal problems in establishing ownership of patents and royalties, points out Joseph Donovan of Prudential Securities. Sometimes, ownership is shared among several people or companies, not all of whom may want to be part of the securitisation.

Together these factors are an obstacle to unusual securitisations. To make them work, issuers may have to pay a big price by securitising only a small fraction of the forecast cash flow, by paying a high interest rate or by providing other safeguards. Last year, pop's Mr Bowie had to persuade his record company to guarantee investors' money in case his music fails to sell as expected before he could complete the $55m securitisation of his future royalties.

Most of this strange paper ends up in the portfolios of pension funds and insurance companies which are willing to bear the risk in exchange for some extra return. When Punch Taverns, a British firm, issued £535m ($880m) of securities in March tied to beer sales and publicans' rental payments at 1,428 pubs, the main buyers were non-British banks. In general, such securities must be held to maturity. Unlike mortgage-backed securities, the price of which changes constantly in response to interest-rate movements, there is no easy way to reckon whether the revenues from future beer sales are worth more today than yesterday. This makes the securities difficult to value and therefore hard to trade.

The possibilities are limited only by investment bankers' imaginations – and the gullibility of the bond-rating agencies. The raters, whose imprimatur is often essential for a security's success, say they are currently refusing favourable ratings to far more proposed issues than they endorse. Despite all the hype about exotic securities, that means that the thriving securitisations are likely to be the duller ones.

Taking Stock: Equity-Based Compensation and the Evolution of Managerial Ownership

We investigate the impact of stock-based compensation, including options and restricted stock, on the ownership of U.S. executives. Equity-based pay spread at explosive rates in the United States during the 1990s. Morgenson (1998) reports that in 1997, the 200 largest U.S. companies had reserved more than 13 percent of their common shares for compensation awards to managers, up from less than seven percent eight years earlier. Institutional investors and shareholder activists have tolerated and even encouraged this diversion of equity to executives, believing that managerial ownership may reduce agency problems. Boards' compensation committees routinely cite the goal of increasing managerial ownership as the rationale for equity-based pay.

Although boards state that they intend stock options and other awards to boost the ownership of managers, executives are not likely to have the same goal. Modern portfolio theory predicts that managers receiving additional stock in their firms should sell these shares or, equivalently, sell other shares they already own, to diversify away the unsystematic risk associated with concentrating wealth in a single asset. This risk is higher for managers than for ordinary investors because executives already have human capital value correlated with firm performance. Whether stock compensation leads to higher managerial ownership therefore appears to be an empirical question related to the strength of managers' diversification impulses.

We study year-to year changes in stock and option ownership in 1993, 1994 and 1995 for 8,516 top managers in 1,646 companies of all sizes, a total of 18,558 person-year observations. We observe inflows of equity to managers' portfolios from new options, new restricted shares, and option exercises, and we compare these with outflows of equity form sales of stock.

Our findings depend crucially on managers' prior ownership. We segment our data into subsamples based on whether an executive owns as many shares as those awarded in new grants of stock options or restricted stock. Executives with low prior ownership exhibit stronger incentives after receiving new options and restricted stock, as we find no evidence of offsetting sales of stock during years with new option awards, and modest evidence of selling after restricted stock awards. For higher-ownership executives, we find active selling during years with new option awards, and in some models, strong evidence of selling after receipt of restricted stock. These sales effectively neutralize much of the incentive impact of high-ownership managers' stock-based pay. For executives who exercise stock options, we find near-total selling of the shares acquired regardless of prior ownership, though this disposal of shares may partly be necessary to raise funds for taxes and to pay the option exercise price.

The data imply that once managers reach a certain ownership level, they actively rebalance their portfolios when boards award equity compensation. We cannot know whether boards expect or condone these sales. Prior research has often overlooked this dynamic aspect of managerial ownership, either treating ownership as exogenous or as a choice variable under the control of the board. Instead, we find that ownership changes endogenously as a function of executive pay and prior ownership. Our findings may interest audiences concerned with managerial incentives, including investors, boards of directors, compensation consultants, and financial analysts. For financial theorists who model the value to executives of stock options and related compensation, our findings call into question the frequent and important assumption that managers cannot hedge the risks of these awards.

Open Market Stock Repurchase Signalling

In the quarter following the stock market crash in October of 1987 there were over 750 new or renewed open market repurchases announced. Following the fall 1997 correction, IBM announced that it was increasing its outstanding share repurchase program by $3.5 billion. Can open market repurchase program share prices, and, if so, how? The answer to former has been documented extensively – for example Vermaelen (1981) finds abnormal price increase of 3.37% after the announcement of a share repurchase. The objective of this paper is to explain how.

How do repurchases increase share prices? The financial press argues that an open market repurchase increases a firm's stock price by increasing its return on equity and its earnings per share. But if a repurchase doesn't affect taxes or the market's information set, then it should have no effect on price. The cost of buying the shared decreases assets by an amount that exactly offsets the reduced number of shares outstanding, leaving value per share (price) unchanged. If the repurchase increases leverage, and if the firm is below its optimal capital structure, then a repurchase should increase the stock price by increasing the interest tax shield. Unfortunately, the tax-shield hypothesis enjoys little empirical support in the literature (for example, see Vermaelen, 1981). There may also be a personal tax advantage to repurchases because they generate a deferred capital gain (to non-tendering shareholders) which is taxed at a lower rate than current cash dividends. But Black and Scholes (1974) argue that firms cannot affect their value by changing their distribution policy. The tax explanations for an increase in firm value are neither theoretically nor empirically attractive.

The academic finance literature has produced three other explanations for the increases in firm value after repurchase announcements: the free-cash-flow hypothesis, the insider-trading option model, and signalling. If firms suffer decreased market value from principal-agent problems, then, following Jensen (1986), a repurchase can increase value by decreasing free cash flow. Howe, He, and Kao (1992) study the relationship between firm value and free cash flow for firms that repurchase via fixed-price offers and do not find support for the principal-agent hypothesis. Since open market repurchases reduce free cash by less than fixed-price offers, we do not expect the principal-agent hypothesis to explain the increase in value attending open market repurchases.

Ikenberry and Vermaelen (1996) (thereafter I&V) model repurchases as an option for asymmetrically informed insiders to buy undervalued shares from uninformed outside shareholders through the facilities of the firm. The announcement period return represents the value of the option to take advantage of the outside shareholders. I&V find support for all three of their model's implications. This paper's signalling model generates two equivalent implications as well as a third implication that is different from I&V's implications, but from a very different set of assumptions. In contrast to the option model where outsiders are unwittingly exploited, the signalling model assumes rational maximization by both parties, and in equilibrium the insiders' information is revealed to the benefit of both inside and outside shareholders.

The last of the academic hypotheses is that repurchases signal information to the market. Vermaelen (1981) and comment and Jarrell (1991) both conclude that the repurchase proportion signals positive information about firm value. The firms themselves, who usually cite undervaluation as the primary motivation for the repurchase, support the signalling hypothesis. Despite the popular support for the hypothesis, there are no published open market signalling models, just models of fixed-price and Dutch auction repurchases.

The advantages of formally modelling an open market repurchase signalling game are that we gain deeper insight into how the signalling mechanism works, and we obtain a set of rigorously derived implications that can be used to test empirically the explanatory power of the hypothesis. The latter is particularly beneficial since past test have not been guided by a formally derived equilibrium and so have only tested one implication: that the market's announcement return is increasing in the repurchase proportion (the presumed signal). This model and its three testable implications are described next.

The intuition for my signalling model is simple because it parallels the effects that a repurchase has on the utility of the firm's controlling shareholder (entrepreneur). I assume that entrepreneurs are risk averse and do not tender into their own offers (evidence in support of the latter is provided in Section I. Part B). These assumptions drive single crossing (the technical condition that assures that the signal is believable). If the entrepreneur refrains from tendering, then the repurchase increases his or her proportionate shareholdings. The repurchase also increases the riskiness of the firm in one of two ways: 1) because the firm pays cash and reduces its holdings of its least risky asset, or 2) because it borrows and increases its leverage. Either way, a purchase increases the amount of risk to which the entrepreneur is exposed. With mean/variance utility, entrepreneurs of firms with greater expected earnings experience less disutility from the added risk and opt to repurchase more than entrepreneurs of firms with lower expected earnings. The equilibrium generates three testable implications, namely that market value should be increasing in: 1) the quantity repurchased, 2) the riskiness of the firm's earnings, and 3) the entrepreneur's stock holdings.

Although simple, the model has a formal equilibrium and yields rich testable implications that are either new or consistent with published empirical studies. Signalling provides an immediate explanation for the abnormal announcement-period returns: the market learns of the firm's greater earnings from the signal and so increases its valuation of the firm. Both Comment and Jarrell (1991) and I&V (1996) report that the repurchase proportion is significantly positively related to the announcement return, which is consistent with the first implication of the model, I&V (1996) also find a significant positive relationship between the firm's total risk and the announcement return, which is consistent with the model's second implication. The third implication is novel and has not been tested before.

Appendix A-13: Reading comprehension test

Name ___________________

Answer the following questions by providing short answers.

STAR-STRUCK

1. Where does the cash from Bowie bonds originate?

____

2. Which American movie is going to be financed through shares issue?

____

3. Which rating agency is rating David Bowie's bonds?

____

4. What is one of the important reasons music-related bonds were not a great success?

____

5. If Civilian Capital makes a lot of money, how will the investors holding their special bonds do?

____

6. What will happen to the value of Bowie's bonds if they are downgraded?

____

BONDS

7. Whose careers are more suitable for securitisation: of an established rocker or of a rapid growing new pop star?

____

8. Who has guaranteed Bowie's securities to investors?

____

9. What are asset-backed securities?

____

10. Why are credit cards and mortgages particularly suitable to be used as assets in asset-backed securities?

____

11. Who are the main buyers of asset-backed securities?

____

TAKING STOCK

12. Why do investors and shareholders tolerate equity-based payment of managers?

13. What could crucially influence the re-sale of awarded stock?

14. What percentage of common shares of the average US company is reserved for equity-based compensation?

15. Why is the risk of holding shares in a company higher for executives of the firm than for an ordinary outside investor?

16. What is the reason corporate boards award equity awards to managers in increasing frequency?

17. Are managers' personal goals likely to be aligned with those of the shareholders?

OPEN MARKET REPURCHASE SIGNALLING

18. Does the stock repurchase increase share price?

19. Why does the price of shares rise after a stock repurchase according to the financial press?

20. What is the main reason for share repurchases according to corporations?

21. By how much does the price of shares rise on average after a share repurchase is announced?

22. What is the relationship between share-repurchase announcements and total risk of a firm?

Appendix A-14: Results: Reading comprehension test

question	correct	largely correct	partly correct	incorrect	not given	irrele-vant	share of at least largely correct answers for each question	facilities	Corrected item-total correlation	Cronbach's Alpha if Item Deleted
1	40%	10%	5%	45%			50%	0.50	-0.065	0.545
2	100%						100%	1.00	0.000	0.498
3	100%						100%	1.00	0.000	0.498
4	65%		15%	20%			65%	0.65	0.062	0.531
5	35%	10%	35%		20%		45%	0.45	0.013	0.462
6	55%	5%	10%	15%	15%		60%	0.60	0.128	0.480
7	80%			20%			80%	0.80	0.224	0.461
8	60%			30%	10%		60%	0.60	0.036	0.472
9	5%	15%	65%	5%	10%		20%	0.20	0.425	0.451
10	70%	5%	5%	10%	10%		75%	0.75	0.116	0.521
11	20%			80%			20%	0.20	0.248	0.465
12	95%	5%					100%	1.00	-0.167	0.504
13	50%	10%	15%	5%	20%		60%	0.60	0.362	0.395
14	55%			45%			55%	0.55	0.134	0.506
15	95%			5%			95%	0.95	0.413	0.444
16	20%	15%	10%	35%	20%		35%	0.35	0.335	0.421
17	80%			5%	15%		80%	0.80	-0.076	0.512
18	100%						100%	1.00	0.000	0.498
19	80%	5%		10%	5%		85%	0.85	-0.194	0.552
20	45%		15%	20%	20%		45%	0.45	0.177	0.489
21	80%			15%	5%		80%	0.80	0.304	0.420
22	50%	5%		40%	5%		55%	0.55	0.238	0.485

Cronbach' Alpha – all 22 items	0.516
Cronbach' Alpha – 17 items last 5 deleted	**0.536**
Cronbach' Alpha – 13 items last 5 deleted + 100%	0.517

Appendix A-15: Background knowledge test

Name________________

Give short answers to the following questions.

1. What is a bond?

___

2. What are tradable shares?

___

3. What are investment-grade bonds?

___

4. What is a bond rating?

___

5. What is a mortgage?

___

6. What is a default on a bond?

___

7. What are royalty payments?

___

8. What are agency problems (or agent-principal problems)?

___

9. Why don’t managers invest only in their own company?

___

10. What is a restricted stock?

___

11. What is an option?

___

12. What is unsystematic risk?

___

13. What is leverage?

___

14. What is a stock repurchase?

___

15. How is an optimal capital structure defined?

___

Appendix A-16: Results: Background knowledge test

question	correct	largely correct	partly correct	incorrect	not given	irrele-vant	share of at least largely correct answers for each question	facilities	Corrected item-total correlation	Cronbach's Alpha if Item Deleted
1	25%	15%	20%	30%	0%	10%	40%	0.40	0.525	0.718
2	65%	10%	5%	5%	15%	0%	75%	0.75	0.244	0.747
3	0%	5%	0%	25%	70%	0%	5%	0.05	0.348	0.743
4	15%	10%	5%	50%	20%	0%	25%	0.25	0.579	0.720
5	55%	0%	15%	5%	25%	0%	55%	0.55	0.159	0.756
6	5%	5%	0%	10%	80%	0%	10%	0.10	0.649	0.711
7	10%	5%	5%	35%	45%	0%	15%	0.15	0.221	0.743
8	30%	15%	15%	10%	30%	0%	45%	0.45	0.220	0.752
9	40%	15%	15%	30%	0%	0%	55%	0.55	-0.301	0.789
10	5%	5%	5%	35%	50%	0%	10%	0.10	0.586	0.724
11	45%	10%	10%	15%	15%	5%	55%	0.55	0.483	0.722
12	30%	5%	30%	15%	20%	0%	35%	0.35	0.505	0.735
13	5%	5%	15%	15%	60%	0%	10%	0.10	0.492	0.731
14	15%	5%	20%	20%	40%	0%	20%	0.20	0.588	0.719
15	10%	5%	15%	25%	45%	0%	15%	0.15	0.569	0.716

Cronbach' Alpha - all 15 items	0.749
Cronbach' Alpha - 14 items, if question 9 is deleted	**0.789**
Cronbach' Alpha - 12 items, if question 13.14.15 is deleted	0.635
Cronbach' Alpha - 12 items, if question 13.14.15 and also 9 is deleted	0.710

Appendix A-17: Post-reading questionnaire

Name ______________________________

Post-reading questionnaire

STAR-STRUCK									
How familiar were you with the topic before you read the text?	Not at all	1	2	3	4	5	6	7	A very great deal
If you were familiar with the topic of the text, how much did this help you answer the questions?	Not at all	1	2	3	4	5	6	7	A very great deal
How easy / difficult was the text?	Very easy	1	2	3	4	5	6	7	Very difficult

Your comment

____

____

BONDS									
How familiar were you with the topic before you read the text?	Not at all	1	2	3	4	5	6	7	A very great deal
If you were familiar with the topic of the text, how much did this help you answer the questions?	Not at all	1	2	3	4	5	6	7	A very great deal
How easy / difficult was the text?	Very easy	1	2	3	4	5	6	7	Very difficult

Your comment

____

____

TAKING STOCK									
How familiar were you with the topic before you read the text?	Not at all	1	2	3	4	5	6	7	A very great deal
If you were familiar with the topic of the text, how much did this help you answer the questions?	Not at all	1	2	3	4	5	6	7	A very great deal
How easy / difficult was the text?	Very easy	1	2	3	4	5	6	7	Very difficult

Your comment

____

____

REPURCHASE SIGNALLING									
How familiar were you with the topic before you read the text?	Not at all	1	2	3	4	5	6	7	A very great deal
If you were familiar with the topic of the text, how much did this help you answer the questions?	Not at all	1	2	3	4	5	6	7	A very great deal
How easy / difficult was the text?	Very easy	1	2	3	4	5	6	7	Very difficult

Your comment

____

____

Appendix A-18: Results: Post-reading questionnaire

		very easy	2	3	4	5	6	very difficult		
	question	1 not at all	2	3	4	5	6	7 a very great deal	9 no answer	mean
SS Topic familiarity	1	50%	15%	15%	10%	10%	0%	0%	0%	2.2
SS Effect of topic familiarity	2	15%	10%	15%	10%	15%	5%	5%	25%	3.5
SS Text difficulty	3	5%	25%	20%	20%	20%	10%	0%	0%	3.6
BRR Topic familiarity	4	55%	25%	0%	5%	15%	0%	0%	0%	2.0
BRR Effect of topic familiarity	5	30%	0%	10%	5%	20%	5%	5%	25%	3.3
BRR Text difficulty	6	0%	10%	10%	5%	25%	30%	20%	0%	5.2
TS Topic familiarity	7	35%	15%	5%	5%	20%	20%	0%	0%	3.2
TS Effect of topic familiarity	8	15%	10%	10%	25%	15%	5%	15%	5%	3.9
TS Text difficulty	9	0%	10%	25%	15%	30%	15%	5%	0%	4.3
OMRS Topic familiarity	10	40%	20%	10%	10%	10%	10%	0%	0%	2.6
OMRS Effect of topic familiarity	11	10%	35%	0%	10%	5%	5%	15%	20%	3.5
OMRS Text difficulty	12	0%	0%	10%	35%	20%	30%	5%	0%	4.9

Appendix A-19: Training thinking aloud and instructions

FLOTATION FEVER

IPOs are making a comeback

You might think that a revival in initial public offerings (IPOs) sales of shares by companies to the public for the first time – would have to wait until the lessons of 1999 and 2000 had faded far into history. Last year, indeed, only $ 45 billion was raised in IPO worldwide, the least in over a decade. How short memories are. A new enthusiasm for new shares is afoot.

In America 29 companies have raised a total of over $ 6 billion this year, compared with three deals producing less than $ 300m by the same point in 2003. Japan has already seen 40 new issues in 2004; more than any other rich country, according to Dealogic, a company which tracks such things. In Europe, even a dull old landline – telecoms firm can offer shares, as Belgacom, Belgium's state-owned telephone company, plans later this month.

Directions to the test-takers:

This is an **investigation into the reading process**.

You're asked to **report what you're thinking** as you read.

Read the passage **up to the first number**, then stop and **think aloud**.

Stop after each number and **describe what you are thinking**

If you don't know, say out loud what you are doing to find out the meanings.

This is **not a test of your proficiency** but an investigation into the reading process.

Here's a passage to try out before we get started with the first text.

Appendix B-1: Grammar test

Circle the correct item (A, B, C or D) to fill the gap. Please stick your code here.

1. "It's highly inflammable, so you ..... better be careful."
   A) had B) would C) should D) must

2. "..... you need some help with the experiment, just tell me."
   A) Will B) Do C) Would D) Should

3. The lecturer informed .....
   A) his decision to us. B) us of his decision. C) his decision for us. D) us his decision.

4. "I am used ..... late at night."
   A) work B) that I work C) to working D) I work

5. ..... of the students has started the course.
   A) Several B) Both C) Neither D) Most

6. "By the time this course finishes ..... a lot about engineering."
   A) I will learn B) I learn C) I will have learnt D) I have learnt

7. "Whatever has happened to the lecturer?"
   "I don't know. He ..... lost."
   A) can have got B) could get C) might get D) may have got

8. If you have read the notes you ..... the answer.
   A) should know B) have known C) shall know D) would know

9. This year the number of candidates who ..... not been worked out.
   A) are successful have B) are successful has C) is successful have D) is successful has

10. My research findings were not ..... to be published.
    A) interesting so B) interesting enough C) enough interesting D) so interesting

11. Coming to study in a foreign country may not be easy at first, but most people eventually enjoy ..... a student in Britain.
    A) being B) be C) been D) to be

12. As a result of his lectures she ..... by this new approach to teaching.
    A) was influenced B) has influenced C) influenced D) had influenced

13. "I ..... yesterday, so I couldn't go to the cinema."
    A) must study B) must have studied C) had to study D) ought to study

14. "If he had known the problems, he ..... the task."
    A) will not have undertaken B) had not undertaken C) should not undertake D) would not have undertaken

15. "I'm going out unless you would prefer me ..... here."
    A) to stay B) will stay C) that I stay D) stay

16. Everyone ..... a difficult course to follow.
    A) believe that statistics is B) believe that statistics are C) believes that statistics are
    D) believes that statistics is

17. It ..... produced the same result whenever we tried it.
    A) ever B) often C) always D) every time

18. "My results are the same ..... yours."
    A) that B) as C) than D) like

19. Until I studied the disease, I ..... its cause.
    A) do not understand B) have not understood C) did not understand D) cannot understand

20. His examination results were not as bad as they ..... been.
    A) need have B) might have C) can have D) must have

21. "Unless I ..... be late."
    A) run, I'll B) don't run, I won't C) don't run, I'll D) run, I won't

22. "You'd better ..... to the doctor next time you feel ill."
    A) to go B) going C) go D) gone

23. A number of the areas the professor ..... in the field of atomic physics.
    A) have specialised in are B) has specialised in is C) have specialised in is
    D) has specialised in are

24. "Have you finished your project?" "Yes, it ..... now."
    A) typed B) is being typed C) types D) typing

25. "Your English is very good." "It should be. I ..... it ever since I started school."
    A) have been learning B) was learning C) had learned D) had been learning

26. "I wonder why he didn't come to class." "He ..... his bus."
    A) can have missed B) could miss C) may have missed D) might miss

27. If only he ..... down the results when he did the experiments!
    A) writes B) had written C) has written D) was writing

28. Caramel is a brown substance ..... by the action of heat on sugar.
    A) form B) forming C) formed D) forms

29. "Would you like some more meat? There's still ..... left."
    A) a little B) little C) a few D) few

30. Not all the students ..... criteria for assessing written work.
    A) understands these B) understands this C) understand these D) understand this

Appendix B-2: Background knowledge test

Please stick your code here.

Give short answers to the following questions. You can answer either in English or Slovene, but don’t just translate.

1. What is a bond?

___

2. What are tradable shares?

___

3. What are investment-grade bonds?

___

4. What is a bond rating?

___

5. What is a mortgage?

___

6. What is a default on a bond?

___

7. What are royalty payments?

___

8. What are agency problems (or agent-principal problems)?

___

9. What is a restricted stock?

___

10. What is an option?

___

11. What is unsystematic risk?

___

12. What is leverage?

___

13. What is a stock repurchase?

___

14. How is an optimal capital structure defined?

___

Appendix B-3: Reading texts 1-3

Star-struck

Share in your idol's success

NEXT month Japanese investors will be able to buy into the careers of five would-be stars. The five – all female and attractive – were chosen from thousands by a firm named Japan Digital Content. It is launching a "talent fund" in which it will sell shares: ultimately, returns will depend on sales of DVDs, calendars and other paraphernalia.

The idea is not entirely new. In 1997 bankers issued bonds backed by future sales of music by David Bowie, a British singer. In exchange for the temporary music rights to 25 of his albums, Mr Bowie received $55m. Investors got an 8 % yield on bonds rated at investment grade. Offerings by other top acts soon followed. David Pullman, the financier who pioneered Bowie bonds, has since issued securities backed by royalties from John Steinbeck's literary estate and from cartoon libraries.

However, equity financing of untested projects is a novelty. The Japanese pop hopefuls are not alone. In mid-November, "Billy Dead", a film to star and be produced by Ethan Hawke, an American actor, became the first movie to issue shares, which will be tradable. Civilian Capital, an online investment bank working with Mr Hawke, is on track to close the $7.9m issue in the next three months, says its president, Barry Poltermann. Filming is due to start in 2004. Investors will have first call on all profits until their money is returned, and then a percentage of earnings – from box-office sales, sequels and merchandise.

DigiCirc, another Californian firm, is also raising money from public investors, this time for music. Investors can buy or lease the rights to songs or albums posted on DigiCirc's website by unknown musicians in return for a piece of future profits.

Is a glamorous new asset-class in the making? Not yet. Moody's, a rating agency, is considering downgrading Bowie bonds. Sales of Mr Bowie's CDs have been soggy, partly because of online piracy; and EMI, the music company that guaranteed the bonds, was itself downgraded by Moody's in March. Indeed, such securitisations have not become as fashionable as many expected. Few musicians own all of the rights to their music; fewer still convince investors that their popularity will last.

Equity investments in unknown talent are even riskier. Sean Kane of Hall Dickler, a New York law firm, thinks that issuing securities backed by pools of intellectual property – from actors, authors and musicians – might solve this problem. Until then, fans-cum-investors may boost their idols' careers and incomes; but they might not be making the cleverest of investments.

Bonds that rock and roll

From cinema tickets to parking fines, almost everything is being securitised. Does it all make sense?

Fancy investing in a security whose payoff depends on how much beer is sold in British pubs? How about a bond to be paid by collections of overdue parking fines in New York City? If you'd prefer, you can purchase the rights to a slice of the revenues from old Italian films, or the amounts raised by selling executive suites in Denver's new stadium, or the royalties earned by pop stars such as David Bowie and Rod Stewart. There is barely a cash flow anywhere, it seems, that cannot be reassembled into a bond-like security that the most conservative of investors might buy.

Putting such strange instruments together has become one of the hottest businesses on Wall Street. New issues of asset-backed securities rose by a quarter last year, to $484 billion. This year's growth rate is expected to be at least as rapid, and the value of new asset-backed securities may for the first time exceed that of traditional corporate bonds. Mostly, that is good news, as securitisation can reduce the cost of borrowing. Yet amid the enthusiasm, the risk may be overlooked. The fact that new sorts of security can be created does not always mean they should be.

It is easy to see why securitisation has become so popular. Issuers gain instant access to money for which they would otherwise have to wait months or years, and they can shed some of the risk that their expected revenues will not materialise. Investment bankers are able to finance their customers without extending loans that tie up large amounts of capital. And investors can hold a new sort of asset, less risky than unsecured bonds, so they can diversify their portfolios and thus reduce their risk.

Those, at least, are the theoretical benefits. Turning them into reality can be tricky. In fact, for all the current excitement about rock-star bonds and the like, it is striking that the vast majority of asset-backed securities are American and involve a few types of assets, such as mortgage and credit-card loans, which generate relatively predictable amounts of cash at predictable times. Much of this year's growth is expected to come from banks securitising loans that are already on their books and from an increase in American-style securitisation in other markets, especially European. Yet even that is by no means certain: two previous attempts to introduce Europeans to asset-backed securities failed, even though American investment banks spent a fortune trying to make them work.

Even traditional sorts of asset-backed securities sometimes spring nasty surprises. Investors in mortgage-backed securities have repeatedly taken hits when borrowers repaid in larger numbers than computer models had predicted. And in the past year, AutoBond Acceptance, a car lender, saw profits drop and GreenTree Financial, which lends on mobile homes, reported a loss after default and pre-payment rates were higher than expected. The securities continued to perform, because investors had purchased only a portion of the expected payments rather than the whole lot.

These failures illustrate how difficult securitisation can be even with relatively straightforward assets. It is harder still to make securities out of such things as music royalties and patent licence fees.

Most traditional securitisations involve cash flows that have been earned, but are not yet received; the main risk is that the cash will flow earlier or later than promised, or not at all.

When lots of similar receivables are bundled together in an asset-backed security, the average rates of default, late payment and pre-payment are predictable, so both issuers and investors can be relatively confident of their ability to value the security properly.

Many of the more exotic asset-backed securities, by contrast, are based not on earned-but-uncollected cash, such as monthly mortgage payments, but on forecasts of future earnings. Henry Morriello, a partner at Kaye Scholer, a law firm in New York, reckons that at least five years of sales data are necessary to have much confidence about future streams of music revenues, which is why established rockers such as Messrs Bowie and Stewart have been able to do deals; it would be much harder to predict royalties for short-lived acts, such as the Spice Girls. Even so, music fans and cinema-goers are fickle, so the past may not be a reliable guide to the future.

Some of the new financial instruments also confront investors with a problem largely absent from traditional asset-backed securities: moral hazard. In this case, that means that the issuer's own actions can have a considerable impact on the value of the securities. Once a pool of mortgages is securitised there is little that the original lender can do to alter its value. The value of PolyGram's recent $650m securitisation of ticket, video and merchandising revenues, however, depends heavily upon how hard the company promotes its upcoming films. There may be moral issues of different sort as well. "What happens to record sales if the star is caught molesting kids?" asks Paul Taylor, of Duff & Phelps, a bond-rating agency.

Odd couples

One reason investors love securities which pool assets is that they limit risk by minimising the loss if something awful happens to any one asset. This is harder to do with royalties, film revenues and patent licences: One rock star's royalty stream differs from another's in a way that two 25-year mortgages rarely do. Although Rod Stewart is reported to have securitised his royalties, that is only half true; he merely received a $15m loan secured against future royalties. Bankers hope to bundle this with other loans to entertainers and fully securitize it, but this is proving hard to do. Just negotiating the simpler deal announced last week took more than a year.

Complicating matters further, it is not always clear whether the issuer of such one-off securities actually owns the cash flow being securitised. There can be huge legal problems in establishing ownership of patents and royalties, points out Joseph Donovan of Prudential Securities. Sometimes, ownership is shared among several people or companies, not all of whom may want to be part of the securitisation.

Together these factors are an obstacle to unusual securitisations. To make them work, issuers may have to pay a big price by securitising only a small fraction of the forecast cash flow, by paying a high interest rate or by providing other safeguards. Last year, pop's Mr Bowie had to persuade his record company to guarantee investors' money in case his music fails to sell as expected before he could complete the $55m securitisation of his future royalties.

Most of this strange paper ends up in the portfolios of pension funds and insurance companies which are willing to bear the risk in exchange for some extra return. When Punch Taverns, a British firm, issued £535m ($880m) of securities in March tied to beer sales and publicans' rental payments at 1,428 pubs, the main buyers were non-British banks. In general, such securities must be held to maturity. Unlike mortgage-backed securities, the price of which changes constantly in response to interest-rate movements, there is no easy way to reckon whether the revenues from future beer sales are worth more today than yesterday. This makes the securities difficult to value and therefore hard to trade.

The possibilities are limited only by investment bankers' imaginations – and the gullibility of the bond-rating agencies. The raters, whose imprimatur is often essential for a security's success, say they are currently refusing favourable ratings to far more proposed issues than they endorse. Despite all the hype about exotic securities, that means that the thriving securitisations are likely to be the duller ones.

Taking Stock: Equity-Based Compensation and the Evolution of Managerial Ownership

We investigate the impact of stock-based compensation, including options and restricted stock, on the ownership of U.S. executives. Equity-based pay spread at explosive rates in the United States during the 1990s. Morgenson (1998) reports that in 1997, the 200 largest U.S. companies had reserved more than 13 percent of their common shares for compensation awards to managers, up from less than seven percent eight years earlier. Institutional investors and shareholder activists have tolerated and even encouraged this diversion of equity to executives, believing that managerial ownership may reduce agency problems. Boards' compensation committees routinely cite the goal of increasing managerial ownership as the rationale for equity-based pay.

Although boards state that they intend stock options and other awards to boost the ownership of managers, executives are not likely to have the same goal. Modern portfolio theory predicts that managers receiving additional stock in their firms should sell these shares or, equivalently, sell other shares they already own, to diversify away the unsystematic risk associated with concentrating wealth in a single asset. This risk is higher for managers than for ordinary investors because executives already have human capital value correlated with firm performance. Whether stock compensation leads to higher managerial ownership therefore appears to be an empirical question related to the strength of managers' diversification impulses.

We study year-to year changes in stock and option ownership in 1993, 1994 and 1995 for 8,516 top managers in 1,646 companies of all sizes, a total of 18,558 person-year observations. We observe inflows of equity to managers' portfolios from new options, new restricted shares, and option exercises, and we compare these with outflows of equity form sales of stock.

Our findings depend crucially on managers' prior ownership. We segment our data into subsamples based on whether an executive owns as many shares as those awarded in new grants of stock options or restricted stock. Executives with low prior ownership exhibit stronger incentives after receiving new options and restricted stock, as we find no evidence of offsetting sales of stock during years with new option awards, and modest evidence of selling after restricted stock awards. For higher-ownership executives, we find active selling during years with new option awards, and in some models, strong evidence of selling after receipt of restricted stock. These sales effectively neutralize much of the incentive impact of high-ownership managers' stock-based pay. For executives who exercise stock options, we find near-total selling of the shares acquired regardless of prior ownership, though this disposal of shares may partly be necessary to raise funds for taxes and to pay the option exercise price.

The data imply that once managers reach a certain ownership level, they actively rebalance their portfolios when boards award equity compensation. We cannot know whether boards expect or condone these sales. Prior research has often overlooked this dynamic aspect of managerial ownership, either treating ownership as exogenous or as a choice variable under the control of the board. Instead, we find that ownership changes endogenously as a function of executive pay and prior ownership. Our findings may interest audiences concerned with managerial incentives, including investors, boards of directors, compensation consultants, and financial analysts. For financial theorists who model the value to executives of stock options and related compensation, our findings call into question the frequent and important assumption that managers cannot hedge the risks of these awards.

Appendix B-4: Reading comprehension test

Please stick your code here.

Answer the following questions by providing short answers. You can answer either in English or Slovene.

STAR-STRUCK

1. Where does the cash from Bowie bonds originate?

___

2. Which American movie is going to be financed through shares issue?

___

3. If Civilian Capital makes a lot of money, how will the investors holding their special bonds do?

___

4. Which rating agency is rating David Bowie's bonds?

___

5. What is one of the important reasons music-related bonds were not a great success?

___

6. What will happen to the value of Bowie's bonds if they are downgraded?

___

BONDS THAT ROCK AND ROLL

7. What are asset-backed securities?

___

8. Why are credit cards and mortgages particularly suitable to be used as assets in asset-backed securities?

___

9. Whose careers are more suitable for securitisation: of an established rocker or of a rapid growing new pop star?

___

10. Who has guaranteed Bowie's securities to investors?

___

11. Who are the main buyers of asset-backed securities?

___

TAKING STOCK

12. What percentage of common shares of the average US company is reserved for equity-based compensation?

---

13. Why do investors and shareholders tolerate equity-based payment of managers?

---

14. Why is the risk of holding shares in a company higher for executives of the firm than for an ordinary outside investor?

---

15. What could crucially influence the re-sale of awarded stock?

---

16. What is the reason corporate boards award equity awards to managers in increasing frequency?

---

17. Are managers' personal goals likely to be aligned with those of the shareholders?

Appendix B-5: Post-reading questionnaire

Please stick your code here.

POST- READING QUESTIONNAIRE

1. Please answer the questions by circling the appropriate number.

Gender	Male **1**	Female **2**					
Do you read articles, magazines or books on **any topic** in English?	Not at all **1**	**2**	**3**	**4**	**5**	**6**	A great deal **7**
Do you read articles, magazines or books **for your coursework** at the university in English?	Not at all **1**	**2**	**3**	**4**	**5**	**6**	A great deal **7**

2. After reading each text, please answer all the questions by circling the appropriate number or by giving a short explanation either in English or Slovene.

STAR-STRUCK

How **difficult** did you find the text?	Very easy **1**	**2**	**3**	**4**	**5**	**6**	Very difficult **7**

Why? What do you attribute the level of difficulty to? Please comment.

____

How **familiar** were you with the topic before you read the text?	Not at all **1**	**2**	**3**	**4**	**5**	**6**	A great deal **7**
How much did this **affect** your reading comprehension?	Not at all **1**	**2**	**3**	**4**	**5**	**6**	A great deal **7**

In what way did your familiarity with the topic or lack of it show in your reading comprehension? Please comment. ____

____

How **interesting** did you find the text?	Not at all **1**	**2**	**3**	**4**	**5**	**6**	A great deal **7**

BONDS THAT ROCK AND ROLL

How **easy / difficult** was the text for you?	Very easy 1 2 3 4 5 6 7 Very difficult

Why? What do you attribute the level of difficulty to? Please comment.

____

How **familiar** were you with the topic before you read the text?	Not at all 1 2 3 4 5 6 7 A great deal
How much did this **affect** your reading comprehension?	Not at all 1 2 3 4 5 6 7 A great deal

In what way did your familiarity or lack of familiarity with the topic show in your reading comprehension? Please comment. ____

____

How **interesting** did you find the article?	Not at all 1 2 3 4 5 6 7 A great deal

TAKING STOCK

How **easy / difficult** was the text for you?	Very easy 1 2 3 4 5 6 7 Very difficult

Why? What do you attribute the level of difficulty to? Please comment.

____

How **familiar** were you with the topic before you read the text?	Not at all 1 2 3 4 5 6 7 A great deal
How much did this **affect** your reading comprehension?	Not at all 1 2 3 4 5 6 7 A great deal

In what way did your familiarity or lack of familiarity with the topic show in your reading comprehension? Please comment. ____

____

How **interesting** did you find the article?	Not at all 1 2 3 4 5 6 7 A great deal

Appendix B-6: Key and expected answers

1. Grammar test: key

1.A, 2.D, 3.B, 4.C, 5.C, 6.C, 7.D, 8.A, 9.B, 10.B, 11.A, 12.A, 13.C, 14.D, 15.A, 16.D, 17.C, 18.B, 19.C, 20.B, 21.A, 22.C, 23.D, 24.B, 25.A, 26.C, 27.B, 28.C, 29.A, 30.C

2. Background knowledge test: expected answers

1. A type security from which an investor can expect coupons and a final repayment of the principal.
2. Shares that are traded on an organised market like a stock exchange.
3. Bonds with a high credit rating as opposed to "junk" or high-yield bonds.
4. A formal assessment of the creditworthiness of a firm by an independent agency.
5. A loan taken out to buy a house or a building with the payments secured by that same house or building.
6. A default occurs when the issuer is unable to pay out the coupon (or part of it) or the principal.
7. Payments by a publisher to a writer, a photographer or a musician for the use of intellectual property rights - percentage of sales.
8. A conflict of interest between principal-owner and agent-manager. The latter has more info. (i.e. asymmetry of info). When a manager becomes a shareholder, they act differently.
9. A stock that can only be sold under certain conditions, e.g. only after some time, or only with the board's approval, etc.
10. A security that gives the holder the right, but not an obligation, to buy or sell an asset at an agreed (predetermined) price on a specific date.
11. Company-specific risk that can be reduced through diversification (also known as diversifiable risk).
12. The level of debt capital compared to equity. (Nivo zadolževanja nasproti financiranju z lastniškim kapitalom)
13. When a company buys back its own shares from the market, thereby reducing the number of shares outstanding (i.e. supply of shares on the market). This increases earnings per share and may increase the value of shares.
14. Proportion of debt vs. equity (relative to assets) at which the weighted average cost of capital (WACC) is minimal, and the share price is maximal.

3. Reading comprehension test: expected answers

Text 1: STAR-STRUCK

1. Temporary music rights to 25 Bowie's albums. Future sales of Bowie's cds.
2. Billy Dead.
3. Very well because they will have first call on profits until their money is returned, and then a percentage of all earnings/ profit.
4. Moody's.
5. Online piracy / downgrading of EMI / musicians don't own the rights to their music / musicians don't convince the investors that their popularity will last.
6. The value will go down.

Text 2: BONDS THAT ROCK AND ROLL

7. Securities that generate cash flows from a particular underlying asset, such as a mortgage, royalty payments, etc.
8. Because the payoffs from these types of assets are relatively well–known (as there is a "paper trail") and relatively easy to forecast – for example, we today know exactly how much we will have to pay for our mortgage during the next 10 years.
9. The career of an established rocker, as 5-year sales data are required.
10. His record company.
11. Big investment and pension funds who have a strong need to diversify as much as they can.

Text 3: TAKING STOCK

12. 13 percent.
13. To reduce agency problem.
14. Because they not only have their money tied in the firm, but also their human capital – e.g., respect of the community, peers, etc. Also, they are more likely to be non-diversified, i.e., they have relatively a lot of their personal wealth tied in the firm as opposed to a well-diversified investor.
15. Prior ownership
16. They believe that managerial ownership will reduce agency problems.
17. No, they will maximise their own utility, and this may differ significantly from the shareholders'.

Appendix B-7: Instructions for administering the test battery

Napotki za izvedbo – v slovenščini

Če so prisotni **tuji študenti**, je treba njihove teste posebej označiti, praviloma sodelujejo le naši študenti. Prosim, če lahko **označite, kdo v vaši skupini ni sodeloval – v seznamu študentov.**

Namen študije: raziskava branja strokovnih besedil v angleškem jeziku
Hvala za sodelovanje!
Anonimna udeležba – študenti dobijo nalepke s kodo (4 nalepke)

Sestavni deli:

1. **Slovnični test (10 min)**
2. **Test poznavanja pojmov (20 min)**
3. **3 besedila & test bralnega razumevanja & vprašalnik**
    - **Star- struck (10 min)**
    - **Bonds that rock and roll (20 min)**
    - **Taking stock (20 min)**

Ad 1) Grammar test

– Študenti nalepijo kodo.

Ad 2) Background knowledge test

– Študenti nalepijo kodo.
– Test preverja, če poznajo določene strokovne pojme.
– Odgovori so lahko tako v angleškem ali slovenskem jeziku.
– Ni dovolj, če besedo samo prevedejo v slovenščino – napišejo naj definicijo ali pojasnijo bistvo.
– Če za pojem še niso slišali, naj to napišejo.

Ad 3) Reading comprehension test & Post reading questionnaire

Najprej **razdelimo študentom test razumevanja in vprašalnik** in jih prosimo, da prilepijo kodo.

Preletimo vprašanja iz vprašalnika, prevedemo jih v slovenščino:

- Kako **težko/lahko** se vam je zdelo besedilo?
- Čemu **pripisujete** zahtevnost besedila?
- **Koliko ste prej poznali temo**, ki jo obravnava besedilo?
- Koliko je vaše poznavanje ali nepoznavanje **vplivalo na vaše branje**?
- Na kakšen način se je vaše poznavanje ali nepoznavanje teme **izrazilo pri bralnem razumevanju**?

Študenti preberejo **prvo besedilo**. Ob tem sproti odgovorijo na 5-6 vprašanj v **testu razumevanja** v slovenščini ali angleščini ter izpolnijo **rubriko z oceno prebranega besedila v vprašalniku.** Odgovorijo naj na vsa vprašanja in pazijo, da vpišejo oceno pod pravi naslov – zaporedje besedil je namreč različno.

Ko preberejo eno besedilo, odgovorijo na vprašanja v testu razumevanja in vprašalniku, učitelj razdeli drugo besedilo in pobere prejšnje. Če se izkaže, da študenti nujno potrebujejo kakšno minuto več, jim jo damo.

Na koncu se študentom **zahvalimo** in jih zaprosimo, če **ohranijo test kot skrivnost** in o vprašanjih ne govorijo kolegom.

Appendix B-8: Training thinking aloud and instructions

FLOTATION FEVER

IPOs are making a comeback

You might think that a revival in initial public offerings (IPOs) sales of shares by companies to the public for the first time – would have to wait until the lessons of 1999 and 2000 had faded far into history. Last year, indeed, only $ 45 billion was raised in IPO worldwide, the least in over a decade. How short memories are. A new enthusiasm for new shares is afoot.

In America 29 companies have raised a total of over $ 6 billion this year, compared with three deals producing less than $ 300m by the same point in 2003. Japan has already seen 40 new issues in 2004; more than any other rich country, according to Dealogic, a company which tracks such things. In Europe, even a dull old landline – telecoms firm can offer shares, as Belgacom, Belgium's state-owned telephone company, plans later this month.

Directions to the test-takers:

This is an **investigation into the reading process**.

You're asked to **report what you're thinking** as you read.

Read the passage **up to the first number**, then stop and **think aloud**.

Stop after each number and **describe what you are thinking**

If you don't know, say out loud what you are doing to find out the meanings.

This is **not a test of your proficiency** but an investigation into the reading process.

Here's a passage to try out before we get started with the first text.

Appendix C-1: Grammar test: Item-total statistics (n=358)

Question	Key	Key code	A	B	C	D	Miss	Facility value
1	A	1	4.2%	3.9%	83.0%	8.4%	0.6%	4.2%
2	D	4	25.7%	23.5%	19.8%	28.8%	2.2%	28.8%
3	B	2	12.6%	65.1%	1.7%	19.8%	0.8%	65.1%
4	C	3	10.1%	23.2%	63.4%	2.8%	0.6%	63.4%
5	C	3	7.8%	2.2%	16.8%	72.6%	0.6%	16.8%
6	C	3	54.7%	3.4%	31.3%	10.3%	0.3%	31.3%
7	D	4	0.0%	6.1%	35.8%	57.5%	0.6%	57.5%
8	A	1	70.9%	2.5%	1.4%	24.9%	0.3%	70.9%
9	B	2	55.6%	37.4%	2.0%	4.5%	0.6%	37.4%
10	B	2	0.0%	76.3%	17.0%	6.4%	0.3%	76.3%
11	A	1	93.0%	0.3%	0.3%	6.4%	0.0%	93.0%
12	A	1	83.2%	11.2%	1.4%	3.1%	1.1%	83.2%
13	C	3	1.4%	10.3%	87.7%	0.6%	0.0%	87.7%
14	D	4	2.2%	5.0%	8.4%	83.5%	0.8%	83.5%
15	A	1	89.7%	0.0%	3.4%	6.7%	0.3%	89.7%
16	D	4	23.7%	5.3%	11.7%	59.2%	0.0%	59.2%
17	C	3	2.2%	11.2%	82.4%	4.2%	0.0%	82.4%
18	B	2	0.6%	90.2%	4.2%	5.0%	0.0%	90.2%
19	C	3	0.8%	24.6%	64.5%	8.7%	1.4%	64.5%
20	B	2	1.1%	91.1%	5.6%	1.1%	1.1%	91.1%
21	A	1	66.8%	2.8%	10.9%	18.7%	0.8%	66.8%
22	C	3	10.6%	4.5%	77.9%	5.6%	1.4%	77.9%
23	D	4	5.3%	29.9%	7.3%	52.8%	4.5%	52.8%
24	B	2	1.7%	89.7%	3.1%	3.9%	1.7%	89.7%
25	A	1	84.6%	2.2%	0.6%	8.9%	3.6%	84.6%
26	C	3	0.3%	4.2%	71.2%	19.8%	4.5%	71.2%
27	B	2	3.1%	77.7%	8.1%	7.0%	4.2%	77.7%
28	C	3	4.5%	5.6%	85.5%	0.3%	4.2%	85.5%
29	A	1	31.6%	31.8%	17.9%	15.1%	3.6%	31.6%
30	C	3	6.4%	11.5%	30.4%	47.5%	4.2%	30.4%

Appendix C-2: Background knowledge: Item-total statistics (n=358)

	Ordinal variables					
	1 Correct	2 Largely correct	3 Partly correct	4 Incorrect	5 Don't know	6 Irrelevant
Q 1	4.7%	23.7%	40.8%	17.3%	6.7%	6.7%
Q 2	49.4%	23.7%	6.1%	8.7%	10.9%	1.1%
Q 3	3.6%	3.1%	1.4%	20.7%	70.1%	1.1%
Q 4	8.7%	2.5%	7.3%	45.8%	34.4%	1.4%
Q 5	37.7%	15.6%	17.3%	9.5%	19.8%	0.0%
Q 6	4.7%	1.7%	1.1%	21.2%	70.4%	0.8%
Q 7	4.5%	0.8%	1.7%	31.3%	61.7%	0.0%
Q 8	32.4%	23.7%	7.8%	13.7%	22.3%	0.0%
Q 9	1.4%	1.1%	7.0%	29.6%	60.3%	0.6%
Q 10	7.5%	8.9%	19.0%	14.2%	16.5%	33.8%
Q 11	6.7%	3.9%	2.8%	59.2%	26.8%	0.6%
Q 12	3.4%	1.1%	6.4%	6.4%	81.3%	1.4%
Q 13	7.0%	0.6%	0.6%	41.9%	49.7%	0.3%
Q 14	0.8%	1.7%	18.2%	36.0%	42.2%	1.1%
Mean	12.33%	8.02%	9.82%	25.40%	40.94%	3.49%

Appendix C-3: Reading comprehension indicators (n=358)

	Ordinal variables					
	1 Correct	2 Largely correct	3 Partly correct	4 Incorrect	5 Don't know	6 Irrelevant
Q 1	40.5%	4.7%	4.5%	42.5%	7.8%	0.0%
Q 2	98.6%	0.0%	0.0%	1.1%	0.3%	0.0%
Q 3	36.0%	38.0%	3.9%	9.2%	12.8%	0.0%
Q 4	89.9%	0.0%	0.0%	8.7%	1.4%	0.0%
Q 5	84.1%	3.6%	3.9%	2.5%	5.9%	0.0%
Q 6	58.9%	0.0%	1.4%	17.0%	22.6%	0.0%
Q 7	4.2%	5.3%	26.5%	45.0%	19.0%	0.0%
Q 8	78.8%	1.1%	3.1%	7.8%	9.2%	0.0%
Q 9	83.2%	0.0%	0.0%	10.9%	5.9%	0.0%
Q 10	72.1%	0.0%	0.0%	18.7%	9.2%	0.0%
Q 11	31.8%	4.7%	0.6%	48.9%	14.0%	0.0%
Q 12	96.9%	0.0%	0.0%	2.2%	0.8%	0.0%
Q 13	90.2%	3.6%	0.6%	3.4%	2.2%	0.0%
Q 14	89.4%	2.5%	1.1%	2.2%	3.4%	1.4%
Q 15	43.9%	3.6%	8.4%	13.4%	29.6%	1.1%
Q 16	5.3%	9.8%	5.6%	33.2%	46.1%	0.0%
Q 17	63.7%	3.4%	2.2%	13.4%	17.0%	0.3%
Mean	62.8%	4.7%	3.6%	16.5%	12.2%	0.2%

Appendix C-4: Post-reading questionnaire: Item response frequencies

A) Gender

Variable/value	1 - male	2 - female	missing
PRQ gender	36.6%	60.1%	3.4%

B) Reading habits

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	missing	Mean	SD
PRQ **RH**q2 - Do you read articles, magazines or books on any topic in English?	4.7%	24.0%	22.6%	14.5%	18.7%	8.1%	5.9%	1.4%	3.67	1.61
PRQ **RH**q3 - Do you read articles, magazines or books for your coursework at the university in English?	5.0%	18.2%	21.2%	26.3%	15.9%	10.3%	1.7%	1.4%	3.69	1.43

C) Text Star Struck

Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	missing	Mean	SD
PRQ **SS**q1 - How difficult did you find the text?	6.1%	21.5%	23.2%	20.1%	18.7%	8.9%	1.4%	0.0%	3.56	1.46

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	missing	Mean	SD
PRQ **SS**q3 - How familiar were you with the topic before you read the text?	32.4%	24.3%	16.8%	12.0%	11.2%	2.5%	0.3%	0.6%	2.54	1.48
PRQ **SS**q4 - How much did this affect your reading comprehension?	10.6%	18.4%	20.9%	20.4%	16.8%	9.2%	3.1%	0.6%	3.54	1.59
PRQ **SS**q6 - How interesting did you find the text?	2.8%	9.5%	14.2%	19.0%	19.8%	20.9%	3.4%	10.3%	4.34	1.50

D) Text Bonds That Rock and Roll

Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	missing	Mean	SD
PRQ **BRR**q1 - How difficult did you find the text?	1.4%	5.9%	11.5%	23.5%	27.9%	22.3%	7.3%	0.3%	4.62	1.41

Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	missing	Mean	SD
PRQ **BRR**q3 - How familiar were you with the topic before you read the text?	30.7%	33.2%	17.9%	8.4%	6.7%	2.2%	0.3%	0.6%	2.35	1.32
PRQ **BRR**q4 - How much did this affect your reading comprehension?	8.7%	20.7%	20.4%	18.2%	16.5%	9.8%	5.0%	0.8%	3.63	1.65
PRQ **BRR**q6 - How interesting did you find the text?	7.5%	15.1%	19.3%	16.2%	17.3%	11.7%	1.7%	11.2%	3.70	1.57

E) Text Taking Stock

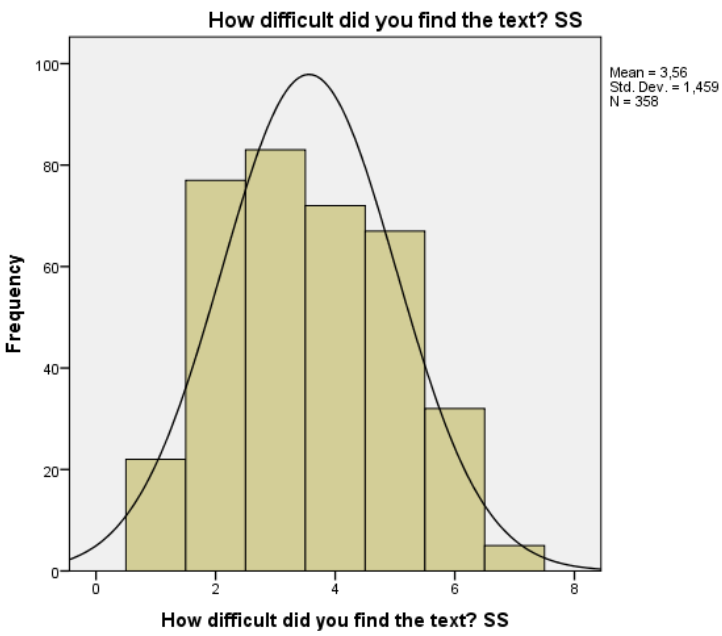
Variable/value	1 Very easy	2	3	4	5	6	7 Very difficult	missing	Mean	SD
PRQ **TS**q1 - How difficult did you find the text?	1.4%	5.9%	11.5%	23.5%	27.9%	22.3%	7.3%	0.3%	4.67	1.36

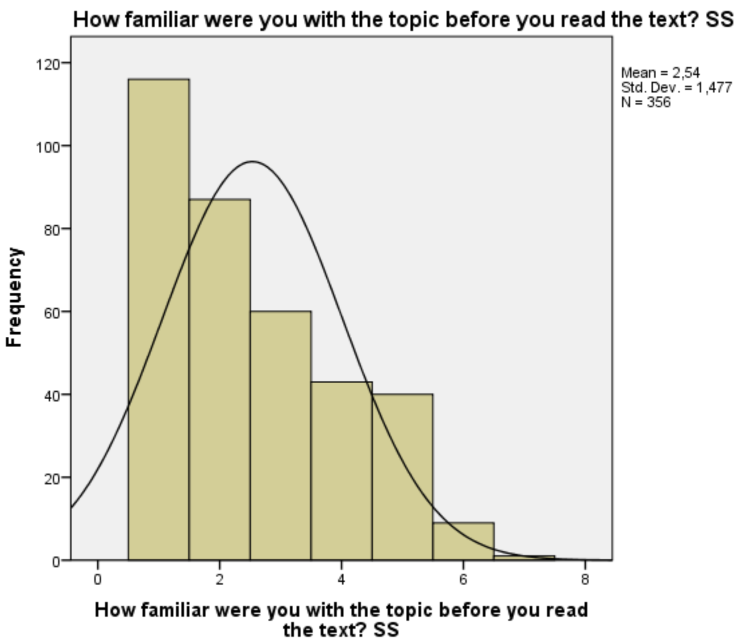
Variable/value	1 Not at all	2	3	4	5	6	7 A great deal	missing	Mean	SD
PRQ **TS**q3 - How familiar were you with the topic before you read the text?	13.1%	20.9%	23.5%	17.9%	15.6%	8.1%	0.6%	0.3%	3.29	1.51
PRQ **TS**q4 - How much did this affect your reading comprehension?	5.6%	15.1%	20.9%	19.3%	20.1%	12.8%	4.7%	1.4%	3.92	1.58
PRQ **TS**q6 - How interesting did you find the text?	9.2%	14.0%	23.5%	19.8%	14.2%	6.1%	2.0%	11.2%	3.47	1.48

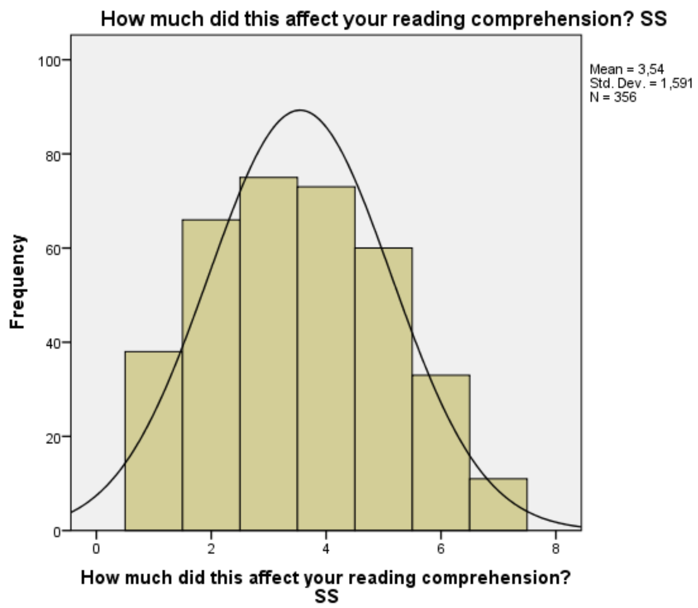
Appendix C-5: Post-reading questionnaire: Star Struck

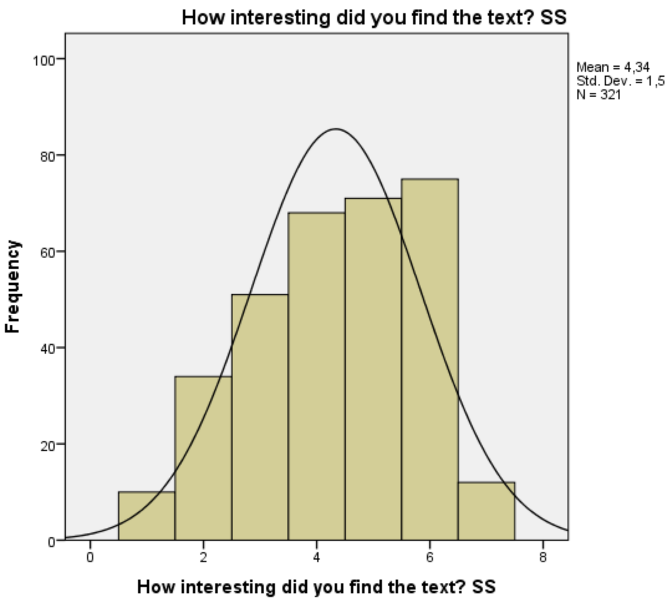
Descriptive Statistics: STAR STRUCK

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
How difficult did you find the text?	358	1	7	3.56	1.459	2.129	.177	.129	-.805	.257
How familiar were you with the topic before you read the text?	356	1	7	2.54	1.477	2.182	.675	.129	-.607	.258
How much did this affect your reading comprehension?	356	1	7	3.54	1.591	2.530	.177	.129	-.782	.258
How interesting did you find the text?	321	1	7	4.34	1.500	2.249	-.314	.136	-.757	.271
Valid N (listwise)	318									




How familiar were you with the topic before you read the text? SS
Mean = 2,54
Std. Dev. = 1,477
N = 356
Frequency
120
100
80
60
40
20
0
0
2
4
6
8
How familiar were you with the topic before you read the text? SS

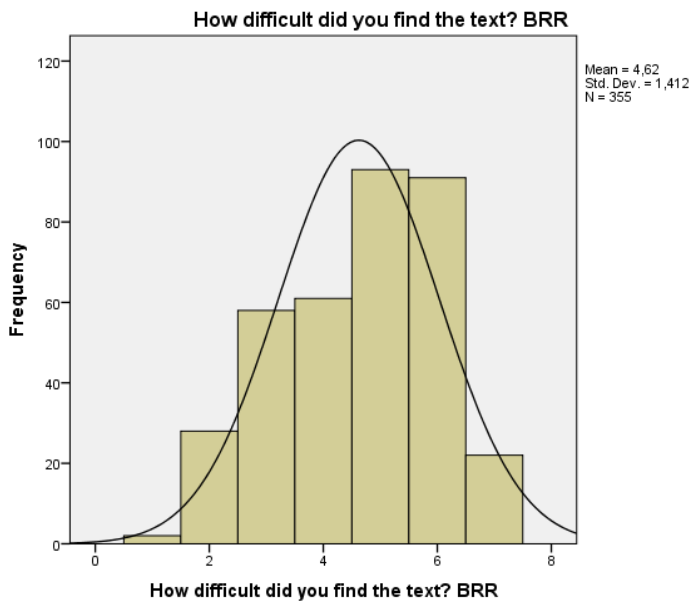

How much did this affect your reading comprehension? SS
Mean = 3,54
Std. Dev. = 1,591
N = 356
Frequency
100
80
60
40
20
0
0
2
4
6
8
How much did this affect your reading comprehension? SS


How interesting did you find the text? SS
Mean = 4,34
Std. Dev. = 1,5
N = 321
Frequency
100
80
60
40
20
0
0
2
4
6
8
How interesting did you find the text? SS

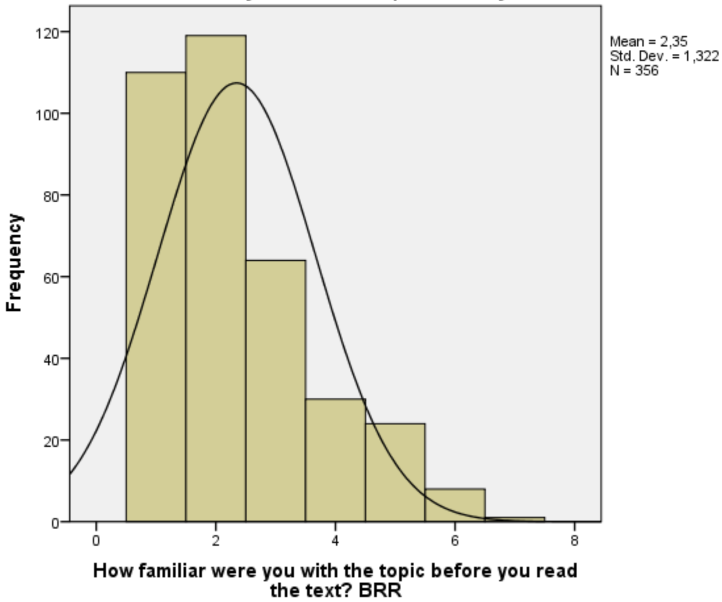
Appendix C-6: Post-reading questionnaire: Bonds That Rock and Roll

Descriptive Statistics: BONDS THAT ROCK AND ROLL

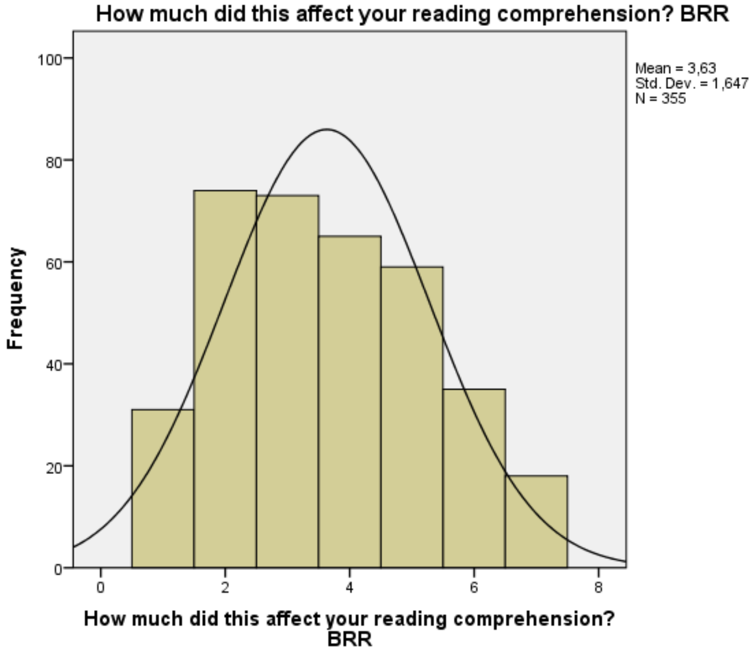
	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
How difficult did you find the text?	355	1	7	4.62	1.412	1.993	-.327	.129	-.774	.258
How familiar were you with the topic before you read the text?	356	1	7	2.35	1.322	1.748	1.007	.129	.432	.258
How much did this affect your reading comprehension?	355	1	7	3.63	1.647	2.714	.257	.129	-.821	.258
How interesting did you find the text?	318	1	7	3.70	1.573	2.474	.027	.137	-.945	.273
Valid N (listwise)	315									

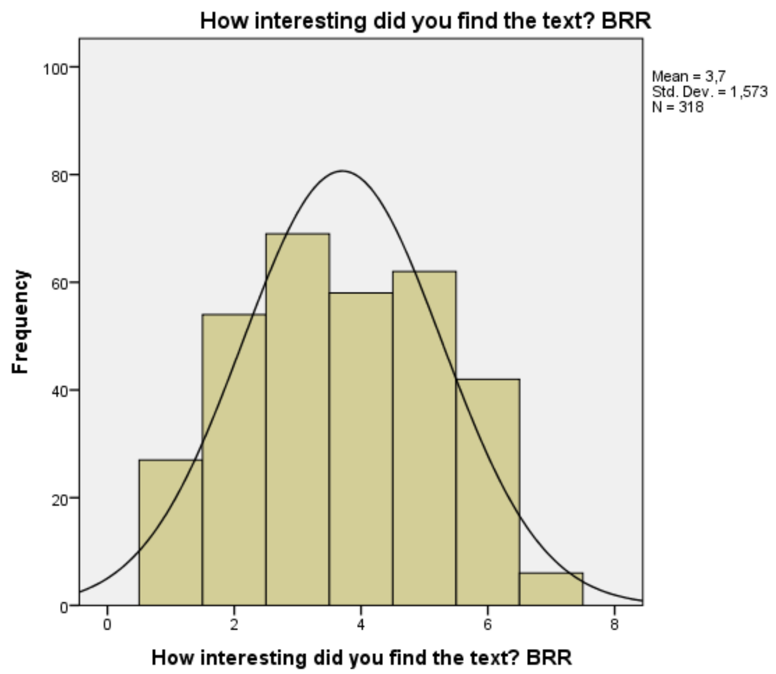


How familiar were you with the topic before you read the text? BRR



How much did this affect your reading comprehension? BRR

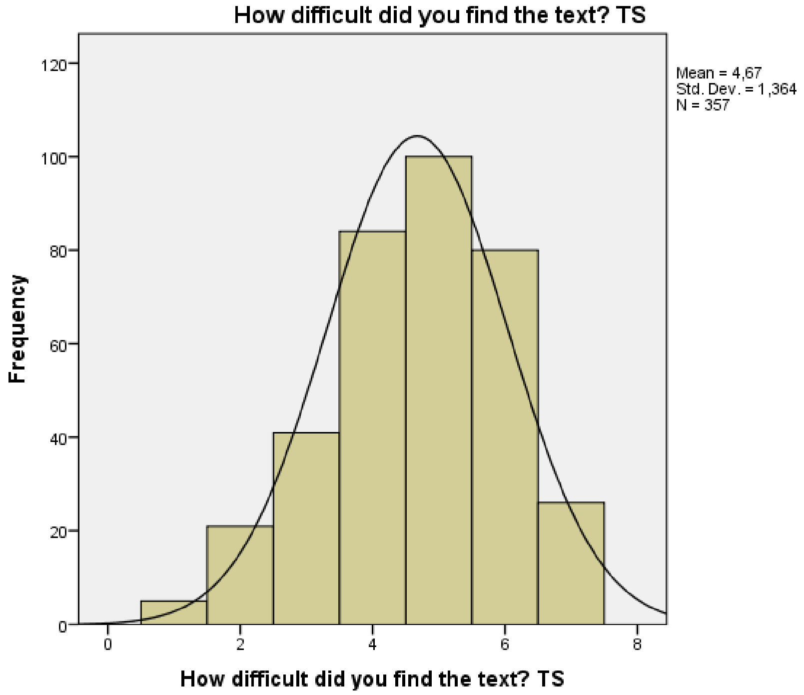


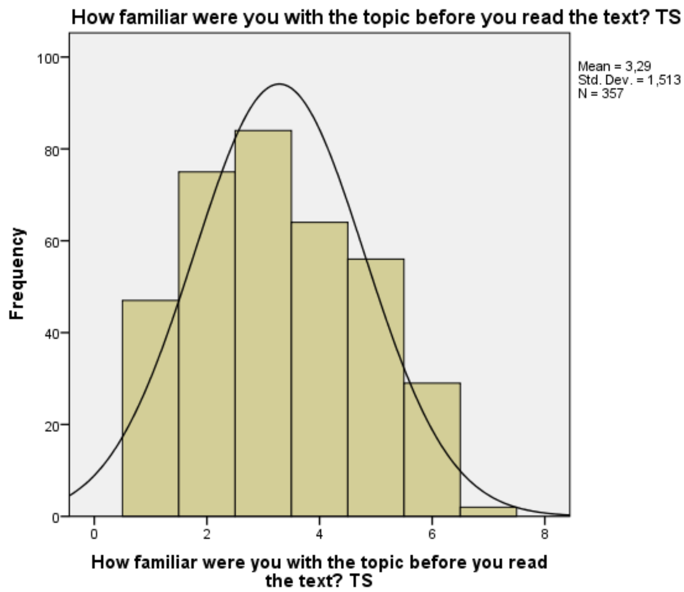

How interesting did you find the text? BRR
Mean = 3,7
Std. Dev. = 1,573
N = 318
Frequency
100
80
60
40
20
0
0
2
4
6
8
How interesting did you find the text? BRR

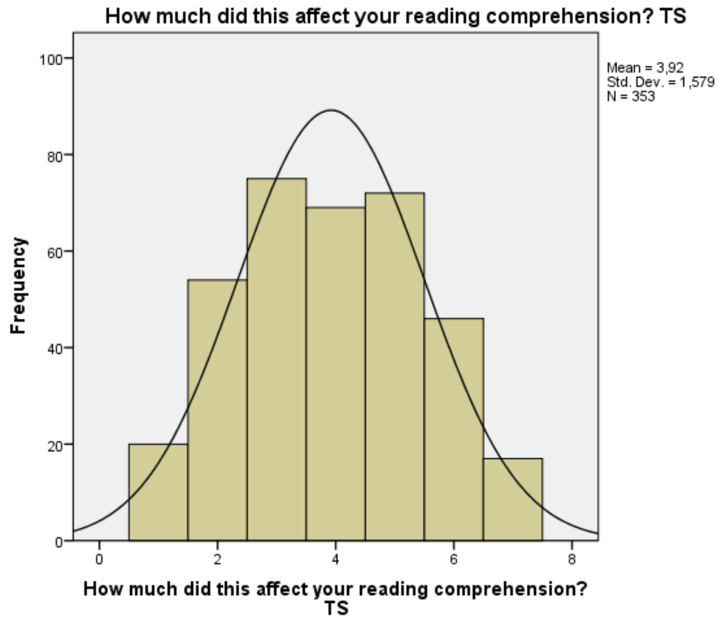
Appendix C-7: Post-reading questionnaire: Taking Stock

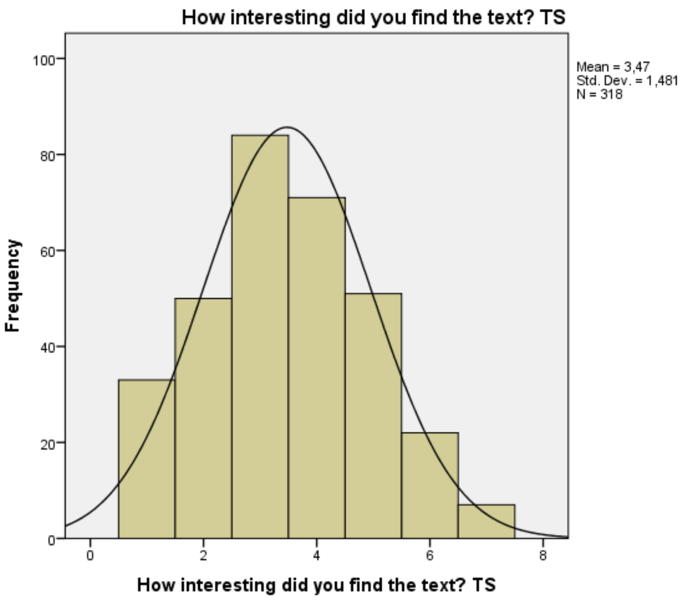
Descriptive Statistics: TAKING STOCK

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
How difficult did you find the text?	357	1	7	4.67	1.364	1.861	-.386	.129	-.279	.257
How familiar were you with the topic before you read the text?	357	1	7	3.29	1.513	2.289	.216	.129	-.875	.257
How much did this affect your reading comprehension?	353	1	7	3.92	1.579	2.494	.044	.130	-.828	.259
How interesting did you find the text?	318	1	7	3.47	1.481	2.193	.178	.137	-.532	.273
Valid N (listwise)	314									




How familiar were you with the topic before you read the text? TS
Mean = 3,29
Std. Dev. = 1,513
N = 357
Frequency
100
80
60
40
20
0
0
2
4
6
8
How familiar were you with the topic before you read the text? TS


How much did this affect your reading comprehension? TS
Mean = 3,92
Std. Dev. = 1,579
N = 353
Frequency
100
80
60
40
20
0
0
2
4
6
8
How much did this affect your reading comprehension? TS


How interesting did you find the text? TS
Mean = 3,47
Std. Dev. = 1,481
N = 318
Frequency
100
80
60
40
20
0
0
2
4
6
8
How interesting did you find the text? TS

Appendix C-8: PRQ results - Perceived topic familiarity & RC for 3 texts

Familiarity SS	**N**	**RC score**
1	116	33.97
2	87	34.34
3	60	34.67
4	43	33.33
5	40	35.98
6	9	32.78
7	1	41.00
Total	356	34.32
Familiarity	N	RC score
1	110	34.45
2	119	33.61
3	64	33.73
4	30	36.03
5	24	36.04
6	8	33.38
7	1	47.00
Total	356	34.29
Familiarity TS	**N**	**RC score**
1	47	31.04
2	75	33.00
3	84	33.25
4	64	35.56
5	56	37.16
6	29	36.90
7	2	39.00
Total	357	34.26

Appendix C-9: Interest & RC for 3 texts

Interest SS	N	RC score
1	10	28.40
2	34	31.59
3	51	33.41
4	68	33.47
5	71	33.69
6	75	36.44
7	12	32.75
Total	321	33.82
Interest BRR	**N**	**RC score**
1	27	29.63
2	54	32.61
3	69	34.32
4	58	33.83
5	62	35.08
6	42	35.07
7	6	35.17
Total	318	33.81
Interest TS	**N**	**RC score**
1	33	32.36
2	50	34.86
3	84	31.89
4	71	34.80
5	51	35.39
6	22	34.18
7	7	34.57
Total	318	33.84

Appendix C-10: Perceived text difficulty & RC for 3 texts

Difficulty SS	N	RC score
1	22	37.45
2	77	36.96
3	83	35.31
4	72	32.97
5	67	32.10
6	32	31.06
7	5	28.60
Total	358	34.25
Difficulty BRR	**N**	**RC score**
1	2	35.00
2	28	35.46
3	58	36.45
4	61	35.80
5	93	34.16
6	91	32.57
7	22	30.77
Total	355	34.31
Difficulty TS	**N**	**RC score**
1	5	34.60
2	21	37.62
3	41	36.29
4	84	35.23
5	100	33.83
6	80	33.58
7	26	28.96
Total	357	34.26

Appendix C-11: Sample think-aloud protocol translated into English

Taking Stock: Equity-Based Compensation and the Evolution of Managerial Ownership

TA: *This has to be connected to shares, aaam their connection with managers and their ownership*

1 We investigate the impact of stock-based compensation, including options and restricted stock, on the ownership of U.S. executives. Equity-based pay spread at explosive rates in the United States during the 1990s. Morgenson (1998) reports that in 1997, the 200 largest U.S. companies had reserved more than 13 percent of their common shares for compensation awards to managers, up from less than seven percent eight years earlier.

TA: *We now study, this article studies what kind of influence had »stock-based« pay - what is that? Let me go on and then make connection with this. What influence had shares on ownership aaam, on ownership of »executives« that is bosses and directors. Aha, this is research, or a report that finds out aaam, that was done among two hundred American companies, and thirteen percent of shares were distributed, actually converted to managers pay. Aaam, this is seven percent less than »eight years earlier«.*

2 Institutional investors and shareholder activists have tolerated and even encouraged this diversion of equity to executives, believing that managerial ownership may reduce agency problems. Boards' compensation committees routinely cite the goal of increasing managerial ownership as the rationale for equity-based pay.

TA: *Aaam, now they say that investors and shareholders have so far tolerated this practice aaam, and why wouldn't they? The reason would be that they believe, that if managers received company shares, this would make smaller the problems with agents, or better between owners and managers. This is actually a problem of the conflict of interest. For instance, managers want to do things their way but the owners want something different: owners can limit managers' pay check, but managers want to have luxury cars, don't they, and this is costly to the owners, and there are more ways how to solve this problem. One committee has found out that managers shares are rising in companies. (WP) Now they will probably say more about this. »Equity-based pay« aaam, is in my opinion this that you are paid according company performance, and you get that back. You are paid according to the success of the* *business operations. This is probably the key to it all – you have a share of the company, that's why you try harder.*

3 Although boards state that they intend stock options and other awards to boost the ownership of managers, executives are not likely to have the same goal. Modern portfolio theory predicts that managers receiving additional stock in their firms should sell these shares or, equivalently, sell other shares they already own, to diversify away the unsystematic risk associated with concentrating wealth in a single asset.

TA: *Aaam, now he talks about how »boards state«, actually, how they use this possibility to give company shares, he now says that managers or »executives » don't have the same goal, they would all like something different. Now he says that modern »portfolio theory«, well some way of thinking, anticipates that managers who get additional company shares could sell them, or could sell other shares they have so that they could spread the »unsystematic, unsystematic« risk. This risk is connected to the number of your investments. So, if you have different investments, you can minimize the risk. Actually, there are two graphs. One is the same continuous line, meaning that risk is always there, no matter what. The other graph shows what happens if you spread your investments, then you reduce your risk to some degree. Let's say, if you have a thousand investments, your risk is the smallest, but if you have half of that, your risk is reduced less. So, where are we? Aha, sales.*

4 This risk is higher for managers than for ordinary investors because executives already have human capital value correlated with firm performance. Whether stock compensation leads to higher managerial ownership therefore appears to be an empirical question related to the strength of managers' diversification impulses.

TA: *They now talk about risk. This is risk which is probably unsystematic, it is higher for normal investors. Because managers have or »executives, already have human capital value correlated with firm performance« I don't know what this could be, »human capital« how is this connected in this case, I don't know that. OK. »Whether stock compensation leads to higher managerial ownership ...« Now he says, that if these shares are given, if managers get company shares and increase their stake in the company, there is an empirical and research question connected to this ability, ability of managers to disperse risk, the risk of the managers.*

5 We study year-to year changes in stock and option ownership in 1993, 1994 and 1995 for 8,516 top managers in 1,646 companies of all sizes, a total of 18,558 person-year observations. We observe inflows of equity to managers' portfolios from new options, new restricted shares, and option exercises, and we compare these with outflows of equity form sales of stock.

TA: *»We study year-to year changes ...« Now he says that they have for some time been dealing with the possibility to get, that managers get shares and increase their shares in companies. Actually, they did some research into that. Let me go back, »stock compensation« – aam, where's that? »Stock compensation« in this case in my opinion it means payment, if I got this right, as a part of pay, which is given in shares, isn't it? This is in my view compensation, although only compensation could mean exchange. Right? But in this case it is compensation for his work. Right? Is that so? OK, he, he. (laughs) We'll I don't quite know. Well, research then. OK, altogether, I'm going to read again the last sentence so that I don't talk too long. Well, there has been quite some research involving eighteen thousand and more people and now what are the results. Aam. Now they compared what kind of influence these shares have on several things. Now he will say more. Aha, the findings now.*

6 Our findings depend crucially on managers' prior ownership. We segment our data into subsamples based on whether an executive owns as many shares as those awarded in new grants of stock options or restricted stock. Executives with low prior ownership exhibit stronger incentives after receiving new options and restricted stock, as we find no evidence of offsetting sales of stock during years with new option awards, and modest evidence of selling after restricted stock awards.

TA: *Now he says that it is very important what kind of share, or what exact share the managers had from before. They divided this sample, their findings into more subsamples. If these managers get more shares than they had before, aaam »options or restricted stock«, aaam »stock option, stock option« is that you can get, let's say managers can get one ..., for instance a manager of Merkur can get today a chance to buy shares at a certain price after two years. This is his privilege, this means that no other small shareholder will be able to buy shares at the same price. So this is a kind of motivation. You don't pay out all in cash, but you give them a chance to buy shares. So in this way you stimulate growth and improvement of the company. I have no idea what is restricted stock, some sort of limited stock, if I translate literally, I really don't* *know, I have never heard this expression. Now he talks about, actually he compares executives with a smaller share and ... he says that they have a stronger influence. »After receiving new options« aha, they would probably want to take advantage when they get »option« or »restricted stock«, »as we find no evidence of offsetting sales of stock during years with new option awards, and modest evidence of selling after« There are no proofs that in line with this new theory that they would sell their shares. As it was mentioned before.*

7 For higher-ownership executives, we find active selling during years with new option awards, and in some models, strong evidence of selling after receipt of restricted stock. These sales effectively neutralize much of the incentive impact of high-ownership managers' stock-based pay.

TA: *Now he talks about the comparison with »high-ownership executives« that is among those that get a bigger company stake, then they realize, well cool he realizes that they are selling their shares. This means the more they get, the more they sell. Also when they receive »restricted stock«, it would be good to know what they are. Aaaam, I have to go back to the previous sentence because I don't know if I understood that correctly, he somehow negates this. »For higher-ownership executives, we find active selling during years with new option awards, and in some models, strong evidence of selling after receipt of restricted stock. These sales effectively neutralize much of the ...« Well, now he says that these sales neutralize, or weaken »an impact« well, the influence of these managers that get larger shares. But I don't know what impact it neutralizes. That's why I read the previous sentence again, but I still don't understand the point. I'm going to read on, I hope I can clarify this later.*

8 For executives who exercise stock options, we find near-total selling of the shares acquired regardless of prior ownership, though this disposal of shares may partly be necessary to raise funds for taxes and to pay the option exercise price.

TA: *Aha, now he says that those managers who aaam, »exercise stock option«, well they use this option, »we find near-total selling of the shares«, they find that they sell almost all shares »near-total selling« aaam, no matter how many shares they owned before. »This disposal of shares may partly be necessary to raise funds for taxes and to pay the option exercise price aaaam, option exercise price«, I don't quite get this, so I'm reading on. Aha, well he lists the reasons, or possible reasons, why they would* *sell this. It »may partly be necessary to raise funds for taxes«, they collect money to pay taxes. »And to pay the option exercise price«, well I said before I don't know what that is. OK, they possibly needed money, so they decided for sale.*

9 The data imply that once managers reach a certain ownership level, they actively rebalance their portfolios when boards award equity compensation. We cannot know whether boards expect or condone these sales.

TA: *»The data imply« I think he suggests that when managers come to a certain level of ownership, they start to sell. Here is »rebalance their portfolios« this means, that they possibly want to invest their money somewhere else, or they want to use it for something else. Now they don't know, mhm, the boards, well »expect or condone these«, well they expect or find this normal, or they even if they expect, it would be the best, yes. »Condone« - keep, probably. »Condone«, so, so, just a moment, I'll read again. »We cannot know whether boards expect aa ne or condone these sales,« possibly they approve of these sales, but I'm not sure.*

10 Prior research has often overlooked this dynamic aspect of managerial ownership, either treating ownership as exogenous or as a choice variable under the control of the board. Instead, we find that ownership changes endogenously as a function of executive pay and prior ownership.

TA: *»Prior research has often overlooked this« aha, now they say that this research is a novelty, because no other research before has focused on this aspect or has taken this perspective. He talks again about this connection between managers' shares and their sales, if they are given shares. Aaam, »Instead, we find that ownership changes ...« now they talk again about this connection that it is very, very important to know whether managers owned a big or small part of shares from before.*

11 Our findings may interest audiences concerned with managerial incentives, including investors, boards of directors, compensation consultants, and financial analysts. For financial theorists who model the value to executives of stock options and related compensation, our findings call into question the frequent and important assumption that managers cannot hedge the risks of these awards.

TA: *Well, these are their findings. Aaam, he says that it would be interesting to those who take part in this company, either as shareholders, investors, boards and financial specialists. »Managerial incentives« aaam, perhaps in this case this applies to these* *managers and their actions. Because it says »Our findings may interest audiences concerned with«, this means that this concerns them. Well, it could be something like that. Now he talks about »for financial theorists«, so for those involved in theory, aaam, now he concludes, well, he gives a conclusion that is based on the fact that managers »cannot hedge aaam, risks of these awards«, I don't know what that would be »hedge«. Aaaam, what could it be? To take risk? No, no. It could be to take risk, or perhaps not. Or perhaps to be held accountable for these awards that they get. Managers cannot be held accountable for what they are going to do with their awards. They were given these shares, now they can do what they want. Or it means something else.*