ASSESSMENT OF TEXTUAL LEVEL EFL READING COMPREHENSION:

A COGNITIVE AND CONTEXTUAL INVESTIGATION

BURCU KAYARKAYA

BOĞAZİÇİ UNIVERSITY

2018

ASSESSMENT OF TEXTUAL LEVEL EFL READING COMPREHENSION:

A COGNITIVE AND CONTEXTUAL INVESTIGATION

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

English Language Education

by

Burcu Kayarkaya

Boğaziçi University

2018

I, Burcu Kayarkaya, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;

- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;

- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature......................................

Date......26.03.2018......

ABSTRACT

Assessment of Textual Level EFL Reading Comprehension:

A Cognitive and Contextual Investigation

The purpose of the study is to investigate whether test takers can understand a text at whole text level upon the completion of a commonly used assessment technique, the multiple choice test. Attaining comprehension at textual level and the formation of macrostructure is critical in interpreting test takers' competence in real life or academic contexts. However, if the multiple choice test technique is ineffective in directing test takers towards tackling reading texts in a way that leads to whole text comprehension, accurate interpretations concerning test takers' future performance cannot be made. To fit the purpose of the study, a test with two reading tasks were created: a multiple choice task and a summary task. Thirty-two undergraduate students from a state university based in İstanbul, Turkey took the test and their performance in macrostructure formation after completing both tasks were quantitatively analyzed. During the completion of both tasks, retrospective verbal protocols were carried out with all participants and their responses during the protocols were categorized to probe into the reading processes participants went through while dealing with both tasks. The findings from these analyses provided substantial evidence towards the inefficacy of the multiple choice test technique in reflecting the ability of test takers' macrostructure formation and thus pointed at a serious threat towards validity of the test technique. The study also put forward distinctive processes that test takers tended to carry out during multiple choice test completion, which were explanatory in the occurrence of such a deficiency of the technique.

ÖZET

Yabancı Dil Olarak İngilizce Bağlamında Metin Seviyesinde

Okuma Becerilerinin Bilişsel ve Bağlamsal Açıdan Değerlendirilmesi

Bu çalışma, sınavlarda sıklıkla kullanılan çoktan seçmeli test tekniğinin öğrencilerin okuma sonrasında metinleri bütünsel olarak anlamalarını ne kadar mümkün kıldığını ölçmek için yapılmıştır. Okuma metninin bütününe hakim olmak ve makro-yapı oluşturmak, öğrencilerin akademik koşullarda ne derece başarılı olacağı konusunda kritik önem taşır. Ancak, eğer çoktan seçmeli test tekniği öğrencileri metinlerin bütününe hakim olacak şekilde okumaya yönlendirmiyorsa, öğrencilerin gelecekteki okuma yetkinlikleri konusunda doğru öngörülerde bulunmak imkansızdır. Bu belirsizliklere ışık tutmak için, araştırmanın amacına uygun olacak şekilde iki okuma sınavı hazırlanmıştır; bunlardan biri çoktan seçmeli sınav, diğeri ise sözlü özet çıkarma sınavıdır. İstanbul'da bulunan bir devlet üniversitesinde, lisans programlarında öğrenim görmekte olan 32 öğrenci bu sınavlara girmiştir ve öğrencilerin sınavlar sonrasında metinlerin makro-yapılarını oluşturma açısından gösterdikleri performans nicel olarak analiz edilmiştir. Her iki sınavın tamamlanması sonrasında öğrencilerden bu sınavları çözerken bilişsel olarak geçtikleri süreçler üzerine retrospektif sözlü değerlendirme yapmaları istenmiştir. Bu incelemelerden çıkan sonuçlar şöyle sıralanabilir: çoktan seçmeli test tekniği, öğrencilerin okuma parçaları için makro-yapı oluşturma becerilerini yansıtmamaktadır ve dolayısıyla bu durum test tekniğinin geçerliğine ilişkin ciddi tehditlere işaret etmektedir. Ayrıca öğrenciler, bir metni çoktan seçmeli soru cevaplamak için okuduklarında, diğer okuma amacına göre (sözlü özet çıkarma) çok farklı bilişsel süreçlerden geçmektedirler.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction to the study

The aim of accumulating research on reading is to explain and extend the current

practices in the field along with the suggestions of potentially effective practices that

enhance the teaching and testing of the skill.

The very idea of teaching a skill for practitioners is to make sure the learner is

able to use the skill with ease in a real life context. That is, the purpose of developing

reading as a skill in learners is to make them be able to cope with written materials

that are encountered in real life, for example in daily life, in a formal situation, or in

an academic context. Weir, Hawkey, Green and Devi (2009) stated in their quotation

below what reading in an academic context requires readers to be able to do:

> In general terms, the reading types covered [in an academic context] are
> expeditious reading, i.e. quick, selective and efficient reading to access
> desired information in a text (scanning, skimming and search reading), and
> careful reading, i.e. processing a text thoroughly with the intention to extract
> complete meanings from presented material. (p.160)

In order to decide whether a learner can do well in such a variety of situations, tests

of different types and sizes are used in practice. Making accurate measurements in

terms of a learner's competency in reading in an academic context is one of the

important measurement concerns for which both small scale and high-stakes

international tests are designed.

Among several frameworks discussing academic reading ability, Enright et al. (2000) put forward in TOEFL 2000 Reading Framework that the items used in sections that measure language skills in the test should be built upon a "purpose-driven" approach emphasizing reading tests should include tasks that direct test takers for "(a) reading to find discrete information, (b) reading for basic comprehension, (c) reading to learn, and (e) reading to integrate information across tasks".

Drawing upon the information above, one requirement an L2 reader in an academic context has to meet is to process a text thoroughly to extract complete meanings from the written material and to read to learn. This skill is the basic of learning from texts in academic environments. However, to what extent the commonly used assessment techniques can be instrumental in making readers reveal such competencies in reading is a question has yet to be answered.

Despite the fact that reading is a widely assessed skill, there is overwhelming discussion in literature claiming that scores obtained in tests measuring reading comprehension actually does not truly represent test takers' understanding of the written material (Keenan et al., 2008; Cutting & Scarborough, 2006). There are several factors that shape reading comprehension process in a test taking condition.

Bachman (1990), discussing method facets as well as trait facets in testing, pointed at a number of factors that affect test performance. He divided method facets into five categories:

(a) Testing environment
(b) Test rubrics
(c) The nature of input
(d) The nature of the expected response
(e) The interaction between input and response

Categories (d) and (e) manifest the importance of what type of response is expected when performing in a test and how the test taker should interact with the text to produce the required response. How a test taker should reveal performance and what he/she has to do is directly related to what the test tasks direct him/her towards doing to successfully accomplish the given task. That puts great emphasis on task formats and whether the task formats provide valid grounds on which true reading ability can be displayed. Therefore, to make sure tests that assess reading ability are evaluating the primary constructs of reading, one of which is the whole text level understanding of the written material, the factors above should be taken into consideration so that they do not affect test takers' performance in a construct-irrelevant way. If test takers' performance is masked by factors other than reading ability, tests cease to be valid indicators of competency.

1.2 Significance of the study

There are quite a few studies in literature that aim at drawing upon the effects of the task format on test performance of test takers. Studies that focus on response format mostly take two or more test formats under investigation and compare results to explore the possible impacts those formats make on test performance of readers. For example, Benson and Crocker (1979) presented empirical evidence towards the effects true-false, multiple choice and matching question formats caused on test takers' performance. Shohamy (1984) examined the effects of multiple choice and open-ended test formats when testing reading comprehension. Kobayashi (2002) compared the impacts of two factors in tests: text organization (well vs badly organized texts) and response format (open-ended, cloze, summary).

In'nami and Koizumi (2009) focused on multiple choice and open-ended question types to study their format effects. Similarly, Rauch and Hartig (2010) studied the dimensionality of reading comprehension as assessed through multiple choice and open-ended response formats.

As will be detailed in the next chapter, studies conducted to discover whether task format has an effect on how reading comprehension is assessed have put forward empirical evidence towards the existence of variance in reader performance. This being the case however, to the researcher's knowledge, there are no studies that focus on the extent to which a certain task type limits the amount of comprehension that could take place otherwise. The present study aims at focusing on the multiple choice format as it is extensively used in reading comprehension tests, and it aims at providing evidence to the fact that tests made up of multiple choice items may not be measuring reading comprehension ability at all levels; i.e. they may be especially weak in assessing a test taker's comprehension of a whole text level. That is to say, a multiple choice reading test may be assessing the comprehension of certain parts of a text but this may not necessarily mean that high scores from such a test reflects a complete understanding of that text. As this is yet another aspect that challenges the validity of multiple choice format in reading assessment, it is important that this question should be probed.

1.3 Research questions

In light of the information presented above, two main research questions have been formulated for the present study. The research questions that we aim to provide answers for in this study are listed below:

RQ1: To what extent can textual level comprehension be attained upon the completion of multiple choice and oral summary reading tasks?

RQ2: How do test takers' reading styles and preferences differ according to multiple choice and oral summary tasks?

1.4 Organization of the thesis

The thesis is comprised of four chapters. Chapter 2 offers a comprehensive literature review on the crucial aspects that need to be discussed before moving further with the research questions: a presentation of reading models, construct validity and cognitive validity in testing, different tasks used in assessing reading, reader purposes and perceptions of tasks, task effect on reading performance. Chapter 3 provides a detailed overview of the participants, instruments, data collection procedures and data analysis methods used to explore the research questions of the present study. Chapter 4 presents the analyses the research questions in terms of the task types and consequences. A detailed discussion of the results in relation to each research question is presented in Chapter 5. Finally, in Chapter 6, a summary of the findings, implications for teaching and testing, limitations of the present study and suggestions for future research are put forth.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The present study aims to investigate whether or how readers' cognitive processes and performance differ while they, learners of English as a foreign language, read texts with different task demands or post-reading activities. To do this, it is necessary to examine reading models and their explanations about what processes readers go through while making meaning out of a written material.

2.2 Models of second language reading

Reading in a foreign language is a complex process with many underlying cognitive components. Examining these cognitive components is necessary to understand how reading and comprehension take place in a reader's mind. Explications on how a reader comprehends texts generally point to a series of steps taken by the reader to eventually arrive at constructing textual meaning or steps taken in a reverse sequence when the reader starts with unfolding the propositions of the text one by one and as a whole and then proceeds with decoding smaller linguistic structures or units. In the bottom-up models that explain reading processes, at the beginning, recognition through the visual system starts at the character level and the characters or letters that are combined form the words. In the next step, words make up the sentences and meaning is derived. This data-driven process is completed through a series of an upward movement, from the bottom to the top (Gough, 1972).

This sequential conceptualization of extracting meaning puts emphasis on decoding, at letter and word level, and failure in decoding makes it impossible to move forward, towards extracting meaning out of idea units that make up the broader, textual understanding.

In the top-down model, a reader goes through a constant process of guessing by sampling the text and predicting what might come next. The reader, then, proceeds in the text to confirm whether what he/she has predicted is realized. This depicts why Goodman (1967) views reading as a "psycholinguistic guessing game", assuming the reader to be sampling, predicting and verifying top-down hypothesis as he/she proceeds in reading. Reading for meaning, in the top-down model, is superior to the mastery of letter/sound relationship and words (Gove, 1983). The reader, being the key in the process, directs the procedure and it is his/her brought-in conceptual or world knowledge that determines whether and to what extent reading is accomplished. That is, "reading is a matter of bringing meaning to print, not extracting meaning from print" (McCormick, 1988).

However, top-down and bottom-up models of reading are criticized as they present an either/or model in a linear manner. The study by Pulido (2004), for example, showed that readers do not always follow a consistent route while reading. The participants in his study shifted between the two models depending on the text's required level of proficiency or prior knowledge. Lee (2009) highlights the importance of topic interest in the selection of cognitive models. Level of proficiency and topic interest are only two of the several factors that are effective in directing readers towards how to read. Therefore, L2 readers tend to follow a bimodal or "interactive" pattern, as will be discussed in this chapter.

Urquhart and Weir (1998) express their concern on the top-down model in the quotation below, casting a doubt on how the top-down process enables readers to take steps in the opposite direction (when compared to the bottom-up process) to construe an overall understanding of a text before dealing with linguistic units that make up the sentences in the text:

> It is virtually impossible to see how a reader can begin by dealing with the text as a whole, then proceed to smaller units of the text, then down to individual sentences, ending with single letters. In fact, the term top-down is deceptive, appearing to offer a neat converse to bottom-up, a converse which in reality does not exist. (p. 42).

Pang (2008) points at the importance of bottom-up models, stating that lower-level processes such as word recognition and lexical-syntactic knowledge are always required in an L2 reading process and that higher-level mechanisms, the ones that operate in top-down models, can only work properly if a reader's language proficiency enables him/her to do so. Similarly, Parry (1991), basing on a two-year longitudinal study on ESL university freshmen, argue that readers, at the stage of vocabulary development, could virtually not activate higher-level cognitive mechanisms and they skip dealing with unknown vocabulary and that at the early stages of language development, readers had no choice other than focusing on tiny segments of language in the hope of generating a broader meaning.

The concern shared in the words of researchers above regarding the top-down process can be summarized in that it is difficult to envision how an L2 reader starts with a most general perspective towards a text at hand and finally concludes the reading activity having attained not only segmental but also textual information.

Given that the ability to handle texts in a top-down manner is mostly associated with readers at higher levels of proficiency, it has yet to explain how readers at lower levels can proceed towards mastery of the written material following a top-down direction.

Trying to alleviate the criticisms levelled against the two models (McCormick, 1988), the interactive model was proposed highlighting the cooperation of top-down and bottom-up models while managing reading. Combining the serial or linear characteristic of the two, interactive model is inclusive of both top-down and bottom-up models (Grabe, 1991) in that it is a blend of the two. As cited in McVee et al., 2005; Rumelhart, 1984 claims that reading is realized both perceptually and cognitively simultaneously and that such distinctions (top-down or bottom-up) are not clear but are quite blurred. Readers go along a continuum of selection of models while reading, changing their focus from linguistic units towards textual clues or the other way round, making use of top-down and bottom-up processes in different quantities and sequences. For example, in a study by Leeser (2007), it was put forward that readers make use of their prior knowledge while dealing with texts. However, when they are confronted with texts with unfamiliar ideas/topic, both processes interact in a complex mechanism or order to facilitate meaning.

The interactive-compensatory model, relying on the same principle as the interactive model, aims at dismissing deficits once they are encountered while reading. It offers an interactive aspect in that readers make use of both bottom-up and top-down processes to decode linguistic information and combine world knowledge with text knowledge.

Compensation is also important in this explanation because the flaws in knowledge source direct readers towards making decisions as to what or which order of processes to follow during reading (Nunan, 1991). Whenever a problem arises while reading, the best problem-solver mechanism is called for, depending on the decision of the reader.

A similar explanation to top-down and bottom-up models describing how reading comprehension is managed is proposed by a number of researchers claiming that reading process is composed of low-level or high-level processes (Grabe, 2009; Kintsch, 2004; Perfetti & Roth, 1980). During low-level processing, readers examine letter/sound relationships to make meaning out of words that comprise sentences that, eventually, help readers arrive at syntactic and sentential meaning. In high-level processing, however, reading activity is maintained at a more global scale, with the aim of constructing mental representations of the written material by gathering and synthesizing information from different propositions and incorporating what has been formed into previous or existing knowledge.

Enright et al. (2000) examined the mechanisms of reading activity in their paper in an effort to define the processing based procedures in reading. The processing perspective aims at describing reading comprehension relying on individual differences that determine or direct reading processes. Main reasons that underlie individual differences in reading ability can be listed as: speed and automaticity of word recognition, thoroughness of word representation knowledge, processing efficiencies in working memory, fluency in syntactic parsing and proposition integration as part of building text comprehension, and the development of an accurate and reasonably complete text model of comprehension.

Carver (1997) in the paper where he develops a model of reading ability building on a number of studies puts forward that reading ability can be defined using a set of competencies in fluency, word recognition accuracy, processing efficiency and reading rate. Works by Geva, Wade-Wooley and Shany (1997) also underline the importance of fluent word recognition, processing efficiency and reading rate while explaining reading ability. The processing perspective, to conclude, regards the construct of reading ability as comprising a set of linguistic and processing variables and thus highlighting the relationship between the competency in the processing skills and reading performance.

Elaborating on the levels and perspectives of reading processes, most discourse comprehension researchers (Grabe, 2009; Kintsch, 1998) accept a two-model account of reading comprehension, in which readers construct both a text model of comprehension and a situation model of interpretation. In the text model, readers figure out the propositions, creating a mental image of the main and supporting ideas of the text. The desired outcome is the matching of the writer's intended message with what readers draw as a conclusion from the text. In the situation model, however, readers build an individual interpretation of the text. Building the interpretation is only possible when a number of factors come into play: reader's prior knowledge, expectations, interest and attitude towards the text. That could explain why the essence of what different readers get from the same text might change depending on the effects of the factors employed.

Whether a reader is more likely to create a text model of comprehension or a situation model of interpretation is linked to several factors, including text type, the reader's level of proficiency, and the purpose for reading.

Grabe (2009) argues that readers are more likely to develop a text model if they have little or inadequate background knowledge towards the topic of the text. Likewise, if readers experience difficulty in understanding the lexical and grammatical aspects of the text such as vocabulary items or linguistic structures, they opt for the situation model, which may provide an inaccurate picture of the text in this case, as the reader relies on background or world knowledge to figure out the essence of the text and that might be quite different from what is intended by the writer of the text.

As mentioned above, one factor that determines whether a reader will construct a text model of comprehension or a situation model of interpretation is what purposes the reader is engaged in the written material for. Carver (1997) classifies reader purposes into two as rauding and reading to learn. While "rauding" a text, readers attempt at getting the gist, or the main points or arguments in a text for general comprehension. However, while "reading to learn", readers try to generate a neat representation of the text mastering both the main and supporting ideas and to integrate what they generate with their domain knowledge. Guthrie and Kirsch (1987) added a third purpose to the list above: search reading or reading to find information, where readers intend to find certain pieces of information by skimming or scanning a text.

Urquhart and Weir (1998), categorize types of reading that serves for different purposes as follows: a) expeditious reading (quick, selective and efficient reading to access desired information in a text - scanning, skimming and search reading), and b) careful reading (processing a text thoroughly with the intention to extract complete meaning from presented material). They further make distinctions between global and local comprehension gains from reading texts.

Global comprehension refers to reaching an understanding of the explicit information available in a text, including main ideas and the links between these ideas, integrating and synthesizing information. The reader is, then, able to build logical relationships between ideas. Local comprehension is more related to an understanding of propositions within the sentence, and it is a process that involves word recognition, lexical access and syntactic parsing and maintaining meaning at the phrase, clause and sentence level (cited in Bax, 2013, Weir & Bax, 2012). A reader, for example, looking for specific information in a text may favor a local expeditious style to reach the necessary information quickly.

However, a reader assigned to read and reflect on what the writer of the text puts forward in a written material may embrace a global careful style that would enable him/her to arrive at a deeper understanding of the text.

Enright et al. (2000) put forward four purposes for reading in L2: a) reading to find information (search reading), b) reading for basic comprehension, c) reading to learn, and d) reading to integrate information across multiple texts. Grabe (2009) built upon Enright's four purposes and added two other purposes: e) reading for quick understanding (skimming), f) reading to evaluate, critique and use information. For some of the purposes listed above (search reading or skimming), a reader does not necessarily form a text or situation model as deriving meaning is not the aim, but for some others (reading to learn or reading for basic comprehension) models of comprehension are required.

Decisions so as to which purpose of reading contributes to the reading activity at hand can be made depending on the type of the task the reader is asked to perform after reading.

The reader purpose perspective by Enright et al. (2000) deals with how a reader approaches a reading text depending on what he/she is supposed to do with the text afterwards, assuming that readers read for a reason in all contexts, be it for an exam purpose or in a real-life context. Successful text comprehension calls for the execution of various strategies during reading (Linderholm & van den Broek, 2002). Adopting the suitable strategy depends on for what reason the reading activity is maintained. For example, reading a leaflet delivered while walking on the street serves a different purpose for the reader than reading a book to write a review. A reader's standard of coherence affects the depth of one's comprehension goals and alters how a text is processed. Requirements of the reading task itself or the goals a reader sets for reading a text change the reading processes a reader goes through while reading. That is, readers shape and reshape their reading behavior or the processes to fit the requirements of the written material.

Brannon (1998 in Linderholm, 2006) investigated the effects of readers' goals on inference generation and memory for expository text. They had college students read for either study or entertainment purposes and found that inference generation during reading is strategic and is affected by the purpose of reading. Linderholm and van den Broek (2002) studied the effects of reading purpose and used the same two reading purposes as in the above study and concluded that readers showed different patterns of processing while dealing with tasks. In another study, Linderholm and Wilde (2010) examined college students' beliefs about comprehension when reading for different purposes: for entertainment and for study. They acknowledged that students engaged in superficial processing while reading for entertainment. However, they went through deeper processing when they were reading for study purposes.

Students stated that their understanding of the text and performance was superior for the study purposes.

It is evident from the above discussion that reading is a complex activity that can take place at different levels, at various depths depending on the purpose of the reader and also task characteristics. Readers with differing proficiency levels and processing mechanisms read texts with varied features, at different difficulty levels and lengths for different purposes, in different contexts. However, as mentioned above, the most important skill in academic environments is the ability to read and understand a text in its entirety with the purpose of learning from it. As mentioned before, this is usually referred to as 'text/situation model formation' (Kintsch, 1998), 'reading to learn' (Enright et al., 2000), 'reading at the whole text level' (Khalifa & Weir, 2009).

The basic principle in this process is the formation of the macrostructure. Macrostructure formation basically defines careful reading processes that are related to whole text comprehension. This process will be detailed below.

2.3 The macrostructure of a text

Understanding the content of a written text can be examined in two levels: at the microstructure level and at the macrostructure level (Kintsch & Kintsch, 2005). The microstructure of a text consists of the explicitly stated information that takes place in the text. To be in control of the text in microstructure level, readers do not necessarily work out the implications or underlying proportions in the text, but they arrive at conclusions by making meaning out of what is explicitly communicated. The microstructure in the sentence "John is going to Mount Everest" provides the reader with the information of where John is going without requiring the need to

make interpretations on part of the reader. However, the probability of enjoying his time there or the necessity to take climbing equipment with him before the journey are just a few of the interpretations a reader who reads that single sentence above can make (Lo, et al., 2016).

A macrostructure consists of the macro propositions in a text. The term "proposition" refers to a conceptual semantic representation that includes the meaning of the sentences in a text that readers store in and recall from their memory. The phrase "the meaning of the sentences" requires a further explanation as it proposes that they do not appear in the exact wording of the original sentence, implying that "meaning" is what is extracted as an end product in a reader's mind after reading. Kintsch (1974, 107) states that "the memory representation of text is a function of the content, but not of the way in which it is expressed".

That is, the reader creates his/her own sentences or ideas after making a sense out of what is written to the extent of the remains of the text in his/her memory. The surface level propositions in a text change form and turn into memory representations to be stored in memory and they are kept in an orderly manner so that the semantic proportions embody the macrostructure of the text (van Dijk & Kintsch, 1983).

Obviously, a reader can make use of both levels when decoding texts, microstructure level of comprehension appearing to be the prerequisite in moving further to the macrostructure level, and macrostructure can be regarded as an aiding tool to reach a deeper meaning. Reaching the interpretations or deeper meanings can be possible at the macrostructure level, where readers' own experience or world knowledge come into play during making meaning out of what has been read.

Therefore, macrostructure is a broad network in which the information being read in a text is assimilated into the reader's existing knowledge (Kintsch, 1994).

How are macrostructures generated? According to the theory of van Dijk (1980), macrostructures are derived from a text through the application of macrorules: less important portions of the text may be deleted, instances may be generalized, and summaries of events may be constructed. The new information presented in the text is combined with what the reader has already in supply and a situational model is produced. The situational model in this case is a group of information that come together to make the whole: the explicit information in the text, how the readers makes meaning out of the explicit information and how these interact with readers' existing knowledge. This "active" process of reading or making meaning comes into existence out of the use of both bottom-up (decoding words to lexicons) and top-down processes (processing sentences upon expectations) (Verhoeven & Perfetti, 2008).

The level of comprehension, that is to say micro or macrostructure formation, is of interest to the field of language testing, as well. Assessing reading ability of a reader is generally carried out by tests that measure how well a reader can answer the questions about a given text. However, it is not always easy to differentiate whether a microstructure or a macrostructure is formed by the reader through the responses elicited for the questions asked. To make this distinction clear, a number of studies were conducted where participants were required to deal with different question types (Deane et al., 2006, Leong et al., 2008). As the purpose was to investigate the reading ability at the microstructure level, the participants were provided with different passages followed by questions that could be answered through the explicitly stated information in the text, such as literal inferencing questions

that did not require the reader to integrate the existing information with the available world knowledge. In other studies, in order to assess reading ability at the macrostructure level, researchers included questions in reading tests that could make readers work out answers by understanding the themes or macro propositions in the texts, for example, elaboration inferencing questions. In their study, Kintsch and Kintsch (2005) claimed that questions targeting the macrostructure are more instrumental in reflecting reading comprehension and ability. The kind of questions researchers used targeted at assessing readers' integration and inferencing ability and tried to reveal whether readers could generate expectations and draw conclusions, which reflected macrostructure formation skill.

It is obvious from the discussion above that macrostructure formation is a high-level reading comprehension stage where the reader forms a text level understanding of the passage s/he is dealing with. It is an important issue to determine whether language assessment tests can tap on the macrostructure formation process or not as this stage defines whether the reader understands the text in its entirety. In a multi-componential language skill as reading, there are many sub-skills (constructs) that operate at different levels from one another and can explain different aspects of the ability such as macrostructure formation or search reading for specific information. In an effort to measure several sub-skills, certain techniques or question types are used in tests to accompany reading texts. The section below will analyze item format issue as an important factor that shapes the reading process.

2.4 The format effect

Although all test formats, in principle, aim at assessing the construct they are designed to tap onto, there is overwhelming data on the fact that different test formats measure different skills or abilities. Cutting and Scarborough (2006) stated that the inferences that are made about how well an individual comprehends the written material vary depending on how it is assessed. As included in Bachman's (1990) framework of test method facets, "the nature of the expected response to the input" that is the test format, and "the relationship between input and response", that is the interaction between the written material and the test format, highlight the importance of taking into consideration the factors that are in constant interplay in testing conditions.

The potential interactions between input (reading texts) and response (post reading tasks) and how a reader interacts with the input to provide the required response has been the center of many studies. There is certainly no best method for testing reading since no single test method can fulfill all the varied purposes for which one might test (Alderson, 2000). Reasons of convenience, practicality and efficiency may come to the forefront while deciding on the most suitable method for assessment. That may, however, cast a doubt upon whether or not the decision is made at the expense of validity.

Enright et al. (2000) also highlight the importance of task demands in their task perspective theory. The theory asserts that the text and task variables are responsible for the good and bad management of reading comprehension tasks. Text and task variables such as the frequency and usage of particular words involved in the task, the complexity of syntax, the amount of text that must be processed, and the amount of time allowed for completing a task may account for much of the variance

in difficulty and task performance on test questions. Kirsch and Mosenthal (1990) also point at the importance of text and task effect in reading texts listing certain important features that could affect comprehension such as the existence or absence of distracting information in the text, the extent to which the correct answer matches the wording of the information in the text, and the concreteness of the information requested. Text and task variables can provide a solid explanation on the factors that cause difficulty for readers while reading texts. Thus, this perspective makes certain that problems or failures in reading performance do not merely stem from the reader, but external factors come into play that make a reader's job easier or harder. The task perspective, on the whole, stresses that it is the text and task that determines whether the reading activity a reader is engaged in will be accomplished successfully. What can be concluded is that there are certain texts and tasks that a certain reader can perform well and there are some others where, the very reader, can do badly. Thus, it is of great importance that texts and tasks be selected carefully and that they tap the underlying cognitive processes in the assessment of reading ability.

Alderson (1990) notes that "answering a test question is likely to involve a variety of interrelated skills, rather than one skill only or even mainly. Even if there are separate skills in the reading process where one could identify one's own reading behavior by a rational analysis process, it appears to be extremely difficult if not impossible to isolate them for the sake of testing or research". The problem with testing reading as Alderson addresses above adds up to the problem of whether or to what extent techniques used for testing reading simulate a real-reader purpose in a real-world condition and that puts greater emphasis on the techniques selected for use.

Benson and Crocker (1979) examined the effect of reading ability and item format (true/false, multiple choice and matching formats) on test performance and concluded that both reading ability and item format significantly affected test performance and that participants outperformed in matching format. Shohamy (1984) researched the impact of various methods used for testing on assessment conclusions. The investigation of multiple choice and open ended question types revealed that particularly the performance of participants with low level of proficiency was affected the most by format used. Keenan et al. (2008) studied how reading comprehension tests vary in the skills they assessed. They compared some of the most popular reading comprehension measures used in research and clinical practice in the United States. Intercorrelations showed that the tests measured different skills. In their paper where they conducted a meta-analysis of test format effects on reading and listening test performance comparing multiple choice and open ended formats, In'nami and Koizumi (2009) stated that multiple choice formats were found to be easier than open ended formats, increasing scores when used in tests. Rauch and Hartig (2010) looked into the effects of multiple choice and open-ended formats in a reading test. It was put forward that open-ended questions were more challenging and time consuming for test takers and that although both question types could measure basic reading abilities such as decoding and word recognition, open ended format could test abilities required to master higher reading processes. Research shows that questions elicit different performance depending on various question attributes such as the type of text, question content, the answer options, the availability of source text while answering (Magliano, Millis, Ozuru & McNamara, 2007).

2.5 Macrostructure formation and format characteristics of multiple choice and summarization tasks

Obviously, some task types are more conducive to assessing text level understanding such as summarization whereas others may only assess local level comprehension processes as in the case with many multiple-choice items. Most language assessment tests, whether in the first or second language, make use of multiple choice items. Therefore, it is important to understand the characteristics of these items types, and scrutinize whether tests formed of multiple choice questions can tap on higher level reading comprehension processes.

2.5.1 Multiple choice technique

Multiple choice technique is one of the most commonly used methods that evaluate reading comprehension. In fact, it is used not only for assessment purposes but also in books and textbooks to teach and enhance comprehension. Although there is not just one solid format of multiple choice question technique, it is usually structured upon a stem - a phrase or a sentence that forms ground to ask a question- that is followed by a number of response alternatives. The reader has to choose the correct or the best one to fit or fulfill what the stem needs. There are several reasons why the multiple choice technique is used frequently in test conditions. First, the "sampling of content is generally superior when compared to other formats" (Haladyna & Downing, 2009). That is, it provides more flexibility to include a bulk of information and more content. Flexibility in design also means that it is possible to address both readers' low-level/high-level processing skills and texts' global/local information sections.

It makes scoring easy and effortless (human raters or machine raters can be in charge), and the answer key objective in that in a carefully planned multiple-choice test, there is a consensus on which alternative is correct and which ones are not. Multiple choice test format offers an economical representation of whether or to what extent the reader comprehends a given text. Take, essay writing using ideas from a text or an open-ended question format, for example. These will cost more on the part of both the reader and the rater as the reader will spend more time deciding on what and how to write and the rater has to make an effort to evaluate what has been written.

The fact that the multiple choice format is commonly used does not make it an ideal one, though. One needs to be aware of both the gains and drawbacks of the techniques used to make dependable assumptions about a reader's performance in reading. While dealing with multiple choice questions, readers usually have to choose the best option from alternatives rather than verbalizing or producing answers themselves. This may mean that the format is inhibiting or limiting what the reader has to say on a given item or point in the text. It also discourages inventive thinking in that it oversimplifies knowledge to fit into one single answer option. What is more, it takes a very skilled test writer to create a well-designed multiple choice test and it is, in this sense, quite time consuming on the part of the test writer. The rater simply does not know why readers respond the way they do. Although scoring can be done objectively (Haladyna, 1994; as cited in Lau et al., 2011) and therefore be reliable, there is the risk of guessing effect (Kurz, 1999), especially if the distractors are not written carefully.

Research and theory suggest that test takers use various strategies to answer multiple-choice items (Anderson et al., 1985; Cohen, 1984), and critics contend that such test-taking strategies are directed by the reader's goal to get an answer that is acceptable rather than to understand the reading selection.

Format effect of question types on reading comprehension has been studied extensively. Most researchers studied the level of difficulty of multiple choice questions posed on readers comparing it to other question types. Shohamy (1984) and Wolf (1991) studied the format effects of multiple choice questions and open ended questions and found that multiple choice formats are easier to deal with for readers. Rupp, Ferne and Choi (2006) provided empirical evidence for the hypothesis that when readers respond to texts followed by multiple choice questions, they go through different processes than they would while reading in non-testing contexts. Rupp et. al. (2006) stated that readers tend to approach the reading text (with multiple choice questions) as a problem solving task, not a comprehension task. That is, readers, as strategic test takers, use a number of techniques to "solve" the problems that appear as questions in test and this makes the activity less similar to real-life reading. That is a big problem when authenticity in test design is taken into consideration, and an issue of test validation.

Cerdan et al. (2009) studied the impact of high-level and low-level questions on superficial level of comprehension and deep level of comprehension. They had a group of participants to be given texts that covered the same textual information and asked them to deal with both high-level questions (answer to which was not explicitly stated but required integration across paragraphs) and low-level questions (answers to which were located in segments of the text and required little inferences).

They divided the participants into two groups; one group read the text first and then answered the questions, the second group was provided with both the text and questions at the same time.

The first group was found to be more successful in integrating ideas in the text and in the analyses made for delayed recall and deep comprehension whereas the second group tended to learn isolated pieces of information. Cerdan et al. (2009) argued that when students read first, they develop a more coherent mental model of the entire text and searched for information more effectively. The research by Cerdan et al. (ibid.) points at the likelihood of readers' following a question-to-text sequence when dealing with multiple choice tasks.

2.5.2 Summarization technique

Though quite rarely, a reading test may contain a task asking test takers to read and summarize all or a part of a written text (Taylor, 2013). Summarizing what is read requires the ability to identify main ideas in the text, integrate them into a text model of reading, and develop a proper situation model of interpretation (Grabe, 2009). In order to understand main ideas, readers need to have a large receptive vocabulary, basic grammar, effective comprehension strategies, strategic processing abilities to maintain a high level of comprehension, and an awareness of discourse structure (Grabe, 2009; Pressley, 2002). The skills also support the fluent reader in establishing the gist of the text. Understanding the main idea in the text helps the reader draw conclusions, evaluate, and critically interpret the content of the text. Summary involves restating the main ideas in a text in the readers' own words and expressions (McNamara, 2007).

In other words, in a successful summary, the reader can differentiate key ideas from supporting ideas and construct logical connections between them.

Kintsch and van Dijk (1978) examine the summarization process in three categories. First, the reader comprehends the text as a coherent whole while removing secondary information. Next, he/she extracts the gist through generalization and finally constructs a new text through generation of recall. As Khalifa and Weir (2009) state "global, careful reading at the highest level requires the reader to understand the micro and macro propositions in a text and how these are interconnected, while integrating new information into a mental model to create a discourse level structure that is appropriate to their purpose".

Robeson (1913, in Taylor, 2013) asserts that there are two characteristics of a good summary. One of them is the inclusion of all that is important and the exclusion of all that is unimportant and not worth mentioning. The other one is the expression of the selected ideas (the ones that are thought to be of primary importance) in a consecutive manner as clearly and briefly as possible.

The summary technique provides a solid picture of how mental processes operate in readers' mind, how they prioritize, construct and organize information as well as the retrieval strategies they use (Bernhardt, 1983). Squire (1993) highlighted the importance of the summarization technique in the following quotation:

> Summarizing, retelling, rephrasing, reprocessing, elaborating, acting out, translating from one medium of communication to another—are vital approaches which require a reader to review, reprocess and recreate the structure of prose. The books we remember, the experiences we best recall are those we have talked or written about. (p. 2)

Brown and Day (1983) stated that summarization is a technique that includes additional and deliberate processing strategies than what are required in comprehension.

As cited in Fisher (2016), Palmer (2003) argued that summarizing is based on a learning process, and the perfect comprehension of a text will be a necessary step in order to teach students to condense information in a new piece of discourse. Summaries can be good indicators of well-developed situation models, when and if readers go beyond the text.

Gil et. al. (2009) showed that summary tasks can be facilitative in that the participants in his study who wrote summaries after reading texts about climate change got higher score points from questionnaires that measured both superficial and in-depth text comprehension (Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, 2009). Summarization is also a preferred teaching task in terms of the washback effect summarizing practice may offer. There are examples in literature that conclusions can be drawn from. Marzec-Stawiarska (2015) studied the influence of summary writing on the development of reading skills in a foreign language. She had the experimental group write summaries of the text they read and the control group engage in test questions including multiple choice tasks. She found that there was a significant difference between the two groups suggesting that summarizing influenced reading ability positively.

In addition, summary technique can enhance long-term retention of information. Mok and Chan (2016) investigated the effectiveness of tests and summary writing tasks in enhancing the long-term retention of students with different levels of test anxiety.

27

They stated that highly test-anxious participants and less test anxious participants in the summarizing condition outperformed those in the test condition.

Summarization is a late developing skill (Brown & Day, 1983) and written products may not show signs of sophistication until well into university years (Garner, 1981). In their paper, Chaka and Booi-Ncetani (2015) provide evidence towards the efficacy of using summary tasks through an oral recall rubric for assessing reading comprehension. They tested their participants (students with poor levels of foreign language) on three formats: a recall task, a summary task and a multiple choice test and found that participants could get a pass score only in comprehension test, suggesting that summary and recall tasks are reliable tools in that poor readers could not manage to accomplish those tasks. The fact that they could get a pass grade from the test task may imply that "proficiency" is not necessarily a factor in the scores participants got from tests, which places doubts on the format used.

Cohen (1994) asked five Portuguese EFL students to summarize English texts to study how participants of different levels of proficiency interacted with texts to produce summaries. Test takers provided verbal report protocols which were analyzed for instances of cognitive processes involved in reading the source text and in writing the summary. It was found that the most successful strategy a user used at least six strategies to her benefit. With respect to reading the source text, she used effectively technical facilitation strategies (underlining discourse markers and words to look up, and circling pronominal referents), clarification and simplification strategies. While writing, she used technical facilitation strategies (providing a detailed answer to include the main ideas) and metacognitive monitoring.

One important consideration in using summaries as a predictor of reading ability is whether they should be in the first language or target language of the reader (Lee, 1986). There are many studies using both patterns. However, it is logical to think that when summaries are produced in target language, written summaries test writing ability, and oral summaries test speaking ability as well as assessing reading. Thus, oral summaries in the first language may be permitted not to overburden readers and have them produce what they have to say without making them spend extra energy and time on thinking how to say it. Another crucial point to keep in mind is the importance of conducting an "immediate" summary task. Otherwise, considerations on memory constraints appear and a reader's performance might also depend upon his/her memory capacity.

With respect to scoring summaries, the practicality of the technique seems to fade away. Considerations of objectivity arise and that puts a strain on reliability issues. Meticulously prepared rubrics, however, can be a solution to circumvent the possibility of subjective rating. The scoring or rating of summary tasks are varied in terms of the methods used. One practical way to assess summaries is to count idea units and ignore structural or meaning relationships (Alderson, 2000). The score readers get is the number of idea units they mention from the text in their recall. The essence of what an "idea unit" may change from design to design, being words or phrases or full sentences. In their paper, Urquhart and Weir (1998) suggest what should be tested and evaluated in the summaries produced by readers is comprehension, but not their interpretation of the texts. However, it is possible that readers produce summaries quite different from each other as they may internalize the propositions differently due to their personal background, prior knowledge, or experience.

In order to remove the disadvantages this may create in scoring summaries, it is important to decide firmly on what should be the common understanding expected out of the text; that is to say, main ideas that should be present in the summary should be determined objectively before scoring.

The above discussion has presented in detail the characteristics of multiple choice and summarization tasks by making clear that summarization tasks can assess macrostructure formation and thus text level comprehension processes whereas multiple choice items may fall short of that capacity. Summarization is a more global and holistic technique that would more closely resemble what readers do when they are reading in real life academic situations. In real-life, it is obvious that readers do not produce summaries after they read texts, but summarization is a task that can easily be used to assess reading to learn processes without much distortion of the cognitive processes. The technique is regarded as a transparent and pure measure of comprehension as there is a direct reader-text interaction and no intervention (by any means of techniques; multiple choice, short answer, etc.). Bensoussan and Kreindler (1990) regarded summary writing as "a whole-text, super-macro-level skill" (p.134).

2.6 Conclusion

In the light of the discussion above, we can conclude that reading is a multi-componential process that is shaped mainly by the purpose of the reader and the task characteristics. Reading can happen at different cognitive levels by achieving different types of understanding of a text; however, academic reading requires careful comprehension of texts for retention and this can only be achieved when whole text level understanding through a successful macrostructure formation occurs. Most reading comprehension tests are designed with the claim that they tap

on a representative collection of reading sub-skills. These tests very frequently utilize multiple choice item types. However, we argue that when there is a set of multiple choice comprehension questions following a reading text, readers tend to focus just on the information that they believe includes the answer for the question.

Thus, whether or not a text is "comprehended" by a reader depends on the extent of the important information (both the main and supporting ideas - a strong representation of the text including writer's intention) that is covered by the questions. That is, particular test designs may direct readers towards a rather "superficial" reading where readers do not have to form a macrostructure in their minds; therefore, do not understand the text in its entirety. That shows how important it is to create tasks that truly make readers focus on and understand whole of the information in a text so that it becomes possible to arrive at the ultimate decision a test writer wants to make: whether the reader possesses the ability to read and understand a text for learning information. The absence of this threatens the validity of any test.

The current study is designed in an effort to shed light on the processes involved in completing a multiple choice test and the level of macrostructure formation at the completion of the test through the comparison of these processes with the ones that occur when test takers read for summarization purposes. The main aim is to scrutinize to what extent multiple choice reading comprehension tests enable readers to form a unified understanding of the text they read as the claim of many reading comprehension tests is to 'assess text comprehension'. The study will also compare reading processes in multiple choice and summarization tasks to underline how limited text processing can be in multiple choice tests. Chapter 3 details the formulated research questions and methodology to achieve this goal.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The purpose of this chapter is to describe the methods used in the investigation of the

research questions of the current study. In doing this, the motivation of the study will

be explained first and then each research question will be handled individually and

one after the other while information about participants, instruments, data collection

procedures and data interpretation methods will be presented.

3.2 Participants

A total of 32 (15 female, 17 male) students were selected for the study. Convenience

sampling was used. In selection of the students, a number of factors were taken into

consideration. As the materials used in the study target learners of English above a

certain level of proficiency (at B2 level), participants were chosen from a pool of

almost 300 students taking an Advanced English mass course at a state university

based in İstanbul, Turkey. When forming the participant group, the grades students

got from the midterm exams were checked and those with scores at or above 80 out

of 100 were shortlisted and a further elimination was made according to the

performance those had in the reading section of the exam. Eventually, the 32

students who were regarded as eligible for the study were asked for consent and all

agreed to participate in the study. The participants signed an informed ethical

consent form (see Appendix A) which was prepared in Turkish, their native language

before they were given the tasks.

3.3 Instruments

3.3.1 Tasks used in the study

To provide answers to the research questions of the present study, two reading tasks were used in the assessment of reading comprehension: a multiple choice task (MC) and an oral summary task (SUMONLY). Whether or not participants who have just completed the MC tasks have readily attained comprehension at textual level will be investigated through an oral summary they will be asked to produce. The data emerging from those summaries (MCSUM) will be compared to that of an independent oral summary task following a second reading text (SUMONLY).

For this aim, a counter-balanced design was used. The study counter-balanced text and task order in a four-way distinction. For example, the first group of participants answered a MC task on Text A, summarized the text (MCSUM), produced verbal protocols (VP) on their reading processes in the first session and then completed an SUMONLY task on Text B as an independent task and produced verbal protocols on their reading processes in the second session. Therefore, there were three tasks, MC test was immediately followed by the summary of the same text (MCSUM), and SUMONLY task done on another text as an independent, summary-only task (SUMONLY). The 32 participants were randomly assigned to the four groups listed in Table 1.

Table 1. The Distribution of Tasks in Four Groups

Group	First session	Second session
Group I N=8	Text A – MC MCSUM VP (How did the participant read for the MC task?)	Text B - SUMONLY VP (How did the participant read the text for the SUMONLY task?)
Group II N=8	Text B – SUMONLY VP (How did the participant read for the SUMONLY task?)	Text A - MC MCSUM VP (How did the participant read for the MC task?)
Group III N=8	Text A – SUMONLY VP (How did the participant read for the SUMONLY task?)	Text B – MC MCSUM VP (How did the participant read for the MC task?)
Group IV N=8	Text B – MC MCSUM VP (How did the participant read for the MC task?)	Text A – SUMONLY VP (How did the participant read for the SUMONLY task?)

3.3.1.1 Multiple choice task

Multiple choice tests are taken from TOEFL preparation materials Collins Practice Tests for the TOEFL Test and Official TOEFL Vol.2. In the original forms, there were 14 multiple choice questions accompanying the texts. Eight questions from each test tapping a variety of reading subskills based on TOEFL test specifications and matching each other in terms of subskills and question type were chosen to follow the texts in the study. Thus, two parallel versions were created. Table 2 shows the different subskill each question both in MC versions of Text A and Text B intends to measure.

Table 2. Cognitive Processes Measured in Questions in Texts

Text A Q1	Negative factual information	Text B Q4
Text A Q2	Text insertion/Cohesion formation	Text B Q5
Text A Q3	Sentence simplification	Text B Q8
Text A Q4	Factual information	Text B Q2
Text A Q5	Inference	Text B Q7
Text A Q6	Rhetorical purpose	Text B Q6
Text A Q7	Reference	Text B Q3
Text A Q8	Vocabulary	Text B Q1

3.3.1.2 Oral summary task

For oral summary reading task (SUMONLY), the same texts, Text A and Text B, were used (see Table 1). The texts with SUMONLY were not accompanied by comprehension questions. The researcher provided a verbal instruction for the participants, informing that what they had to do with the text was to read to it to summarize the text orally after reading it. The participants were free to summarize the text in their L1, Turkish, or in L2, English, depending on their preference.

3.3.1.3 Reading texts of the tasks

Test of English as a Foreign Language, TOEFL, is a strong representative of the kinds of MC EFL tests students encounter in college and university settings. Therefore, the reading texts, carefully selected to fit the purpose of the study, were chosen from the books "Collins Practice Tests for the TOEFL Test" and "Official TOEFL Vol.2". To minimize domain knowledge effect, texts that are found to be of less visited or lesser known topics were chosen for the participants taking part in the study. In other words, texts on topics that are not proximate to fields of study participants were pursuing. To ensure comparability in terms of vocabulary, topic, language use and level, cohesion, coherence, syntactic simplicity, narrativity, genre and interest; automatic text analysis tools (see Table 3) were used along with ideas

and suggestions from expert judgement. The experts consulted were three lecturers working at the Modern Languages Department at a state university, based in İstanbul, Turkey and they had more than at least five years of experience in teaching undergraduate students academic reading and writing. The experts were asked to read and rate the texts in terms of difficulty, relevance for the participant profile and reader interest. Both the automatic text analysis tools available online and expert judgement returned acceptable results in terms of text comparability of the two parallel versions. The copies of both texts can be found in Appendix B.

Table 3. Text Analysis Reports

	Text A	Text B
Genre	Expository	Expository
Title	Which hand did they use?	Training the brain
Word Count	729	734
Readability		
Flesch Reading Ease Score (0-100)	64.1	48.3
Flesch-Kinkaid Grade Level (0-12)	9.8	11.3
Coh-metrix L2 Readability	10.5	10.8
SMOG	9.3	10.8
Vocabulary Complexity		
K1+K2 Word Percentage	82.03	82.97
AWL Percentage	6.31	6.40
Type and Token Ratio	0.43	0.51
Lexical Density	0.57	0.58
Text Features %		
Narrativity	20	27
Syntactic Simplicity	44	59
Word Concreteness	66	57
Referential Cohesion	29	5
Deep Cohesion	39	42

3.3.1.4 Retrospective Verbal Protocols

A retrospective verbal protocol (VP) was used during data collection to investigate the participants' cognitive processes they went through during reading for both tasks (MC and SUMONLY). Therefore, once the summarization process (either after MC task or in SUMONLY condition) was over, the participants were asked to reflect on their reading behavior, explaining how they read the texts with multiple choice tasks and whether or not reading for such a purpose affected their reading style. During the trial sessions, it was observed that some test takers were not openly expressive in verbalizing how they managed the tasks and that they had difficulty in explaining the processes clearly. It was, therefore, decided that a few questions serving as prompts be used to help the participants express with ease how they handled the task. The prompt questions were asked in participants' L1, in Turkish. Table 4 lists the English versions of the questions directed to the participants in verbal protocol at the end of the MC test.

Table 4. Questions Directed to Participants at the end of the MC Task

1. Now that you have finished a multiple choice reading task, what can you say to describe your reading experience?
2. Does your reading style change when you read for such a multiple choice task?
3. Did you read in a linear or segmental manner?
4. Is it fair to say that you read all sentences and paragraphs in the text?
5. Did you follow a question-to-text or a text-to-question order?

After summarization, a similar retrospective verbal protocol was carried out. The participants were asked to comment on their reading behavior, explaining how they read the text with the SUMONLY task to be done later on and whether or not reading for such a purpose affected their reading style (see Table 5).

By asking the questions, the researcher intended to have participants comment on their reading styles and preferences.

Table 5. Questions Directed to Participants at the end of the SUMONLY Task

1. Now that you have finished an oral summary task, what can you say to describe your reading experience?
2. Does your reading style change for such an oral summary task?
3. Did you read in a linear or segmental manner?
4. Is it fair to say that you read all sentences and paragraphs in the text?

3.4 Scoring summaries

When scoring summaries, consulting experts in the field is usually necessary to decide on what makes an acceptable summary of a given text. The researcher worked with the three instructors at YTU who had been consulted for expert judgement. That collaboration was favored in that the instructors were already familiar with the texts used in the study as they read them several times to rate the texts from several aspects to make decisions about their appropriateness for the participant profile. A further consultation session with the three instructors was called for to get their opinions about the ideas and parts of the texts that are essential to be included in a proper summary of the texts. After having them make their own lists of those ideas and parts, a final session was organized with the participation of all instructors to compare a total of four lists (including the researcher's) and to discuss discrepancies and finally reach a consensus for the statements to be looked for in the summaries formed by the participants. As discussed in 2.4.2, summaries can provide a solid ground where readers, when assigned a summarization task, are expected to attain comprehension at whole-text level making a coherent representation of the text by taking out secondary information (Kintsch & van Dijk, 1978) and focusing on

primary information. Primary information that takes place in a reading material is the body of main ideas in the paragraphs. Drawing on this, seven statements (seven propositions with primary information) for Text A and Text B separately were put together to form the rubric for summaries to be used in the assessment of the success of macrostructure formation (see Appendix C). Those seven statements in both rubrics were basically the main ideas of each paragraph in the texts that were identified by the experts and the researcher as important.

While scoring, the researcher paid attention to what extent the statements in the participants' summary matched the ones in the rubric. The participants were allowed to use either their target or native language depending on their preference not to hinder the summarization process with verbalization challenges.

While scoring the summaries produced by participants, intra-rater reliability was taken into consideration due to time and practicality issues. Each summary the participants created were listened and scored twice by the researcher: during the sessions and upon completion of all sessions, time interval between the two scorings was 15 days. While listening to each participant's summary, what participants included in their productions rather than how successfully they verbalized it was examined. During the first listening, the researcher used the rubric for summaries as a checklist and ticked off the sentences in the rubric when (and if) they are included in a participant's summary. At the end, the ticked sentences were counted and a score for the summary was calculated. During the second listening of the video recorded verbal protocols, the summary rubric was again used by the researcher to determine which sentences in the rubric were uttered by the participant while s/he was summarizing the texts.

The scores obtained from the first and second listening of the summaries in two conditions, MCSUM and SUMONLY, were compared and intra-rater agreement (See Appendix D) was found to be .76 and .73, respectively (Cohen's kappa). The final scores were then turned into percentages.

3.5 Coding verbal protocols

As described in the procedural steps followed during data collection, the participants, having completed the two tasks, were asked to express how they read the texts in both tasks through retrospective verbal protocols.

For the coding of the verbal protocols, several coding schemes were examined (Ünaldı, 2004, Cohen & Upton, 2007, Weir et al., 2009, Lim, 2014). As a result, a list of reading operations that fit the purpose of the current study was composed. Table 6 shows the nine reading operations adapted from several studies with two more operations (marked by asterisk) that emerged from the present verbal protocol data added to generate a customized list of operations the participants in the study stated they went through during the completion of the tasks.

Table 6. Coding Scheme for Retrospective Verbal Protocols

_____

No.   Reading operations readers carry out during reading. The participant…

RO1	reads the text carefully before attempting the task[1].
RO2	reads the text expeditiously to have a general idea before attempting the task[2].
RO3	reads expeditiously to find a relevant part in text that is thought to include the answer[3].
RO4*	reads carefully only the selected part(s) of a text that is thought to be relevant to the question.
RO5	have read the whole text from beginning to the end (carefully)[4].
RO6	during reading, reads a part of the text more than once to understand it[5].
RO7	reads to make connections between paragraphs or parts[6].
RO8	tries to understand how the text is organized; how the ideas and details connect each other[7].
RO9	reads to get the main ideas of the paragraphs or parts (and remember them)[8].
RO10*	pays further attention to the introduction and conclusion paragraphs of a text, as they are thought to include the main idea of the whole text.
RO11	follows a question-to-text sequence (and matches the keywords in questions an text)[9].

The verbal protocols were coded and scored twice by the researcher; firstly, during data collection sessions and secondly, after all sessions were finished. The VPs following the two tasks in the study were coded using the coding scheme above, identifying each reading operation as test takers verbalized them. In other words, there were two tasks in the study, a summary following the MC task, the MCSUM and the summary task that was included in the study as the baseline task, the SUMONLY. Therefore, two verbal protocols by the participants were coded. Following this, a second coder (an instructor working at a state university in İstanbul who also took part in expert judgment process) used the coding scheme both for the MC and SUMONLY tasks while watching the video-recorded sessions.

The two coders' results were compared and inter-rater agreement on the reading operations in MC task was found to be 81% and for the SUMONLY task, the

[1] Cohen and Upton, 2007
[2] Weir et al., 2009
[3] Weir et al., 2009
[4] Weir et al., 2009
[5] Ünaldı, 2004
[6] Weir et al., 2009
[7] Weir et al., 2009
[8] Lim, 2014
[9] Weir et al., 2009

41

agreement on the reading processes were calculated as 73% (Appendix D).
Inconsistent results were fixed and agreement upon the finalized lists of reading
operations that each participant mentioned were formed (See Appendix E for the
finalized tables for both tasks).

3.6 Data analysis

The methods of statistical analysis used in the data analysis in the present study will
be given according to the research questions. Table 7 shows the research questions of
the study.

Table 7. An Overview of the Research Questions of the Study

RQ1: To what extent can textual level comprehension be attained upon the completion of multiple
choice and oral summary tasks?

RQ2: How do test takers' reading styles and preferences differ according to multiple choice and oral
summary tasks?

3.6.1 Research question 1

To provide an answer for this question, the researcher used the summary rubric to
count how many of the sentences in the rubric were produced by the participants
while they summarized in two conditions. By doing so, the extent to which the
macrostructure of the text had been successfully formed by the participant in both
conditions was assessed. The two scores for the two summaries were then compared
to find out whether the scores pointed to a statistical difference in the formation of
macrostructures.

For the analysis of scores, a paired samples t-test was used to test the null hypothesis
that there is no difference between the means of two summaries (MCSUM and

SUMONLY). In addition, a two-way analysis of variance (ANOVA) was used to investigate whether the texts or the tasks accounted for the main variance.

3.6.2 Research question 2

In order to establish an answer for this question, the participants were asked to describe how they had read the text in both conditions. To assist them in the retrospective verbalizing process, the questions given in Tables 4 and 5 were asked by the researcher. Their answers were matched with the reading operations in the coding scheme detailed in Table 6. The analysis of the responses, i.e. codes per participant enabled the researcher to visualize what kind of a reading route the participants followed when they were given different tasks. The reading operations in both conditions were compared. For this, the researcher and an expert coded the operations separately (Appendix D). The percentage of the exact agreement in the coding was taken as the reliability of the coding (81% for MC condition, 73% for summary condition). The two coders then formed a final list of operations (see Appendix E) going back to the video recordings and resolving their disagreement. The comparison of reading operations in both tasks are expected to form a basis on which the potential differences in reading outcomes on part of readers can be explained. The statistical significance of the difference was tested by one-sample paired t-test.

3.7 Conclusion

This chapter aimed to explain the procedures and methods used in the exploration of the research questions of the present study. The relevant statistical procedures that were used to investigate each research question were also described. In Chapter Four, the results arising from the analyses of the research questions will be detailed.

CHAPTER 4

RESULTS

4.1 Introduction

As discussed in the previous chapters, the current study was conducted to examine whether the participants could attain comprehension at whole-text level and eventually form the macrostructure of texts they had read upon the completion of two reading tasks: multiple choice and oral summary. The comparison of the success in macrostructure construction in both conditions gave us ideas about the extent of "reading comprehension" that can take place in a multiple choice reading test. In addition, with the retrospective verbal protocols carried out to investigate the reading processes or operations the participants went through, an attempt was made to shed light on the differences that are expected to emerge between the operations executed during reading for the MC task and the SUMONLY task. In this chapter, the analyses of the statistical data obtained for the research questions formulated to investigate the purposes above will be presented in detail.

4.2 Results

4.2.1 Research question 1: To what extent can textual level comprehension be attained upon the completion of multiple choice and oral summary tasks?

The mean scores participants obtained (converted into percentages) from the tests used in the current study were compared to find out whether or not there were significant differences between the scores of all tests. The results showed that participants received the highest score in the SUMONLY task and that overall

performance differences among the MC and SUMONLY task scores were minimal.

The lowest scores were observed in MCSUM, when the participants were asked to

summarize the text they had just read for the MC task, indicating that the participants

received the lowest scores when they summarized the text upon the completion of

the MC task. Table 8 shows the distribution of scores in tests. Tables 9 and 10

present the descriptive statistics for the methods and texts used in the study.

Table 8. Descriptive Statistics

	MC	MCSUM	SUMONLY
M	63.35	48.6	65.59
Median	62.5	42.8	71.4
SD	19.71	22.64	21.13
Skewness	-.2	.36	.06
Kurtosis	-.7	-.64	-.93

Table 9. Descriptive Statistics for the Methods

	MC	MCSUM	SUMONLY
Mean	63.35	48.6	65.59
Median	62.5	42.8	71.4
Std. Deviation	19.71	22.64	21.13
Skewness	-.2	.36	.06
Kurtosis	-.7	-.64	-.93

Table 10. Descriptive Statistics for the Texts

	Text A	Text B
Mean	56.05	62.32
Median	57.1	66.95
Std. Deviation	22.74	21.62
Skewness	.16	-.19

Kurtosis	-.4	-1.1

The full list of scores participants got from MC, MCSUM and SUMONLY are given in Appendix H. However, the RQ1 can be answered through an analysis of scores participants received from MCSUM and SUMONLY.

Therefore, tentatively we concluded that in both cases, the participants could form the macrostructure of the texts but to differing success levels. In order to find out whether the differences between the scores from different task conditions were statistically significant, one-way analysis of variance, with LSD post-hoc test was performed (see Table 11). Here the important difference to observe was between MCSUM and SUMONLY conditions as we were interested in seeing whether MC test tasks were an impediment to the forming of the macrostructure of a text.

Table 11. One-way Analysis of Variance

(I) Method	(J) Method	Mean dif.(I-J)	Std. E.	Sig.	95% CI Lower bound	Upper bound
MC	MCSUM	14.75	5.22	.006	4.37	25.122
	SUMONLY	-2.23	5.22	.670	-12.6	8.14
MCSUM	MC	-14.75	5.22	.006	-25.12	-4.37
	SUMONLY	-16.98	5.22	.002	-27.35	-6.6
SUMONLY	MC	2.23	5.22	.670	-8.14	12.6
	MCSUM	16.98	5.22	.002	6.6	27.35

There was a statistically significant difference between MCSUM and SUMONLY scores of the participants (p=.002). Post-hoc comparisons using LSD indicated that the mean score for MCSUM (M=48.61, SD=22.64) was significantly different from SUMONLY score (M=65.59, SD=21.14). That is, the participants performed better in the macrostructures they formed for the texts when they read for SUMONLY task

and in the MCSUM, the macrostructures they produced of the texts were significantly less successful in comparison to the performance they revealed for the SUMONLY task.

To assure the pure effect of the task or method used in the study in the shaping of such a result, two-way analysis of variance was performed taking score as the dependent variable, the method (task type) and the text as categorical independent variables. The values of kurtosis and skewness (See Tables 9 and 10) were within the accepted levels (i.e., -2/+2), suggesting that the scores based on the method and text are normally distributed. Table 12 shows that the difference between the mean scores of the two tasks was merely a result of the method and that the result was not affected by the texts used in the study or by text and method interactions; the effect for method was significant (F=6.24, p=.003).

Table 12. Two-way ANOVA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7987.41[a]	5	1597.48	3.66	.005
Intercept	336291.53	1	336291.53	771.15	.000
Method	5449.64	2	2724.82	6.24	.003

Text	944.38	1	944.38	2.16	.145
MethodText	1593.39	2	796.69	1.82	.167
Error	39247.85	90	436.08		
Total	383526.81	96			
Corrected Total	47235.27	95			

Thus, for RQ1 it is found that comprehending a text at whole-text level is less likely to take place when test takers read the text for MC task completion. However, when test takers read for a SUMONLY task, there is evidence towards better whole-text level comprehension. This means that although the participants in this study could form successful macrostructures of the text they read under more normal reading conditions, which we attempted to simulate with the SUMONLY task, the participants could not perform equally in comprehending the text in its entirety when they read for the MC task.

4.2.2 Research question 2: How do test takers' reading styles and preferences differ according to the tasks (MC and SUMONLY) they are assigned?

Given that the current study attempted to find evidence towards differences in textual level comprehension of texts when participants deal with them for different purposes, it would be instrumental to make observations on how the participants of the study maintained the reading activity in the two tasks in determining which reading operations (if any) would appear distinctly at whole-text level comprehension. Comparisons of and the differences between the reading styles and preferences in the two tasks are expected to show us how limited text processing can be in multiple choice tests. In an effort to do this, retrospective verbal protocols were used to probe into the reading operations the participants in the study went through while they were reading for the completion of the two tasks: MC and SUMONLY.

The responses the participants (N=32) in the study provided for the prompt questions asked (See Table 4) were examined and eight reading operations for MC task completion were found to be operationalized. Table 13 lists the reading operations that participants declared they went through during reading for the MC task.

Table 13. Reading Operations the Participants Stated They Went Through during the MC Task

	f	%
RO3 follows a question-to-text sequence, matching the keywords in text and questions	30	93.7%
RO6 have read the whole text from the beginning to the end carefully	12	37.5%
RO5 reads carefully only the selected part(s) of the text that is thought to be relevant to the question	9	28.1%
RO7 during reading, reads a part of the text more than once to understand it	8	25%
RO1 reads the text carefully first, before attempting the task	7	21.8%
RO2 reads the text expeditiously to have a general idea before attempting the task	6	18.7%
RO4 reads expeditiously to find a relevant part that is thought to include the answer	6	18.7%
RO9 tries to understand how the text is organized, how the ideas and details connect each other	1	3.1%

When these reading operations are examined, RO3 "follows a question-to-text sequence, matching the keywords in text and question" stands out to be the most frequently stated reading operation by participants (93.7%).

37.5% of the participants maintained that they eventually "have read the whole text from the beginning to the end carefully (RO6)" and almost a quarter of them described their reading process saying that they "read carefully only the selected part(s) of the text that is thought to be relevant to the question (RO5)" and that they "read a part of the text more than once to understand it during reading (RO7)."

The following are examples of some of the reading operations (translated into English) as expressed by participants after reading for the MC task:

Generally, when I am reading for multiple choice, I look at the questions first because I am familiar to [the test technique]. Looking at the questions, I can understand where the answer is. (P8, referring to RO3)

I cannot say that I read the whole sentences and paragraphs here. I read the questions and looked at the text to find [the answers]. To find them, I read it [the part/section] carefully. (P23, referring to RO5)

Table 14 below lists the reading operations participants in the study stated they carried out during the SUMONLY task. The participants responded to the questions (See Table 5) serving as prompts during retrospective verbal protocols. Their responses revealed that eight reading operations were carried out by the participants.

Table 14. Reading Operations the Participants Stated They Executed during the SUMONLY Task

	f	%
RO1 reads the text carefully, before attempting the task	31	96.8%
RO6 have read the whole text from the beginning to the end carefully	31	96.8%
RO10 reads to get the main ideas and remember them	12	37.5%
RO7 during reading, reads a part of the text more than once to understand it	11	34%
RO8 reads to make connections between paragraphs or parts	10	31.2%
RO11 pays further attention to introduction and conclusion paragraphs as they are thought to include the main idea	7	21.8%
RO9 tries to understand how the text is organized, how the ideas and details connect each other	5	15.6%
RO2 reads the text expeditiously to have a general idea before attempting the task	1	3.1%

It is evident that two reading operations came to the forefront while the participants dealt with SUMONLY task: RO1, "reads the text carefully first, before attempting the task" and RO6, "have read the text from the beginning to the end carefully", both with an overwhelmingly high percent, 96.8%. The second most frequently utilized reading operation was RO10, "reads to get the main ideas and remember them", which was reported by 37.5% of the participants. Almost a third of the participants

said they "read a part of the text more than once to understand it during reading, (RO7)". The following are examples of some of the reading operations (translated into English) that participants said they went through while reading for the SUMONLY task:

> As I wanted to make a summary after reading, I wanted to remember it (the text). Especially the main ideas of the parts… I wanted to understand them. (P1, referring to RO10)

> Yes, I read the text in full. Every sentence and every part [of it]. (P15, referring to RO6)

When we compare the differences in the percentages each reading operation (RO) received in the analysis, we can see that there are important differences in the reading operations MC and SUMONLY tasks activated. (See Figure 1)
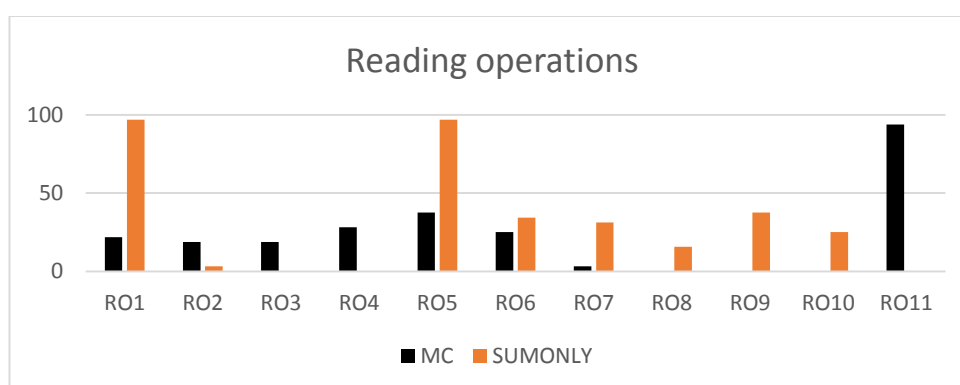
Figure 1 Reading operations by task

A further analysis to investigate whether there are meaningful differences between the means by which each reading operation was reported in the verbal protocols, we performed a paired-sample t-test, and Table 15 below shows the means of reading operations of MC and SUMONLY tasks.

A paired-samples t-test indicated that results were significantly different in the means of every reading operation executed for MC and SUMONLY tasks except for RO6 (See Table 16). The effect size (d= 0.27) for this analysis was found to be low.

Table 15. Paired Samples Statistics – Reading Operations

		M	N	Std. D.	Std. E. M.
Pair 1	RO1MC	.22	32	.42	.07
	RO1SUM	.97	32	.17	.03
Pair 2	RO2MC	.19	32	.39	.07
	RO2SUM	.03	32	.17	.03
Pair 3	RO3MC	.19	32	.39	.07
	RO3SUM	.00	32	.00	.00
Pair 4	RO4MC	.28	32	.45	.08
	RO4SUM	.00	32	.00	.00
Pair 5	RO5MC	.38	32	.49	.08
	RO5SUM	.97	32	.17	.03
Pair 6	RO6MC	.25	32	.44	.07
	RO6SUM	.34	32	.48	.08
Pair 7	RO7MC	.03	32	.17	.03
	RO7SUM	.31	32	.47	.08
Pair 8	RO8MC	.00	32	.00	.00
	RO8SUM	.16	32	.39	.06
Pair 9	RO9MC	.00	32	.00	.00
	RO9SUM	.38	32	.49	.08
Pair 10	RO10MC	.00	32	.00	.00
	RO10SUM	.25	32	.44	.07
Pair 11	RO11MC	.94	32	.24	.04
	RO11SUM	.00	32	.00	.00

Table 16. Paired Samples T-test – Reading Operations

| | | M | SD | Std. E. M. | 95% CI of the D. | | df | Sig. (2-tailed) |
					Lower	Upper		
Pair 1	RO1MC-RO1SUM	.75	.44	.07	-.9	-.59	31	.000
Pair 2	RO2MC-RO2SUM	.15	.36	.06	.02	.28	31	.023
Pair 3	RO3MC-RO3SUM	.18	.39	.07	.04	.33	31	.012
Pair 4	RO4MC-RO4SUM	.28	.45	.08	.11	.44	31	.002
Pair 5	RO5MC-RO5SUM	-.59	.49	.08	-.77	-.41	31	.000
Pair 6	RO6MC-RO6SUM	-.09	.64	.11	-.32	.13	31	.414
Pair 7	RO7MC-RO7SUM	-.28	.45	.08	-.44	-.11	31	.002
Pair 8	RO8MC-RO8SUM	-.15	.36	.06	-.28	-.02	31	.023
Pair 9	RO9MC-RO9SUM	-.37	.49	.08	-.55	-.19	31	.000
Pair 10	RO10MC-RO10SUM	-.25	.44	.07	-.4	-.09	31	.003
Pair 11	RO11MC-RO11SUM	.93	.24	.04	.84	1.02	31	.000

Thus, it can be put forward for RQ2 that reading a text for the MC task and the SUMONLY task required that the participants maintain the reading activity in a different way, making use of different reading operations. We conclude that the MC task directed the participants towards the execution of local level reading operations, which served only for the answering of the questions to differing success levels but did not contribute to the formation of the macrostructure to an optimum level. Thus, the strategic reading operations activated for the MC task did not prove to be successful in whole-text comprehension.

4.3 Conclusion

Below are the conclusions drawn for RQ1 and RQ2:

For the RQ1: As measured by the summarization performance of the participants after reading for the MC task, we cannot say that reading a text for MC task completion guarantees whole-text comprehension to the optimum level, which can be achieved in another condition.

Reading a text for SUMONLY task enables comprehension at textual level; when reading to summarize a text, test takers are more likely to form optimum macrostructures of it.

For the RQ2: The participants read the text for the MC task utilizing mostly MC-specific operations or multiple choice test taking strategies such as following a question-to-text sequence and reading carefully only the selected part(s) of the text that is thought to be relevant to the question.

The participants read the text for SUMONLY task following different reading operations. They engaged in the text to understand it completely and to extract main ideas.

CHAPTER 5

DISCUSSION

5.1 Introduction

Macrostructure formation is possible when and if comprehension at textual level is

attained upon reading a text. Comprehending a text at textual level requires that

certain reading operations be executed. The analyses that were conducted in an

attempt to provide evidence towards the formation of the macrostructure of the texts

used in the current study were reported in Chapter 4 along with the reading

operations participants stated they carried out during accomplishing the two tasks. A

comprehensive discussion of the results derived from the analyses will be presented

in this chapter.

5.2 Research question 1

Research question 1: Can textual level comprehension be attained upon the

completion of MC and SUMONLY tasks?

To provide an answer for RQ1, the participants' summaries in MCSUM and

SUMONLY were compared in terms of the formation of macrostructure of the text

upon reading. The comparison of mean scores of MCSUM and SUMONLY showed

that there was a significant difference between the scores with a mean difference of

-16.98 (see Table 1). That is, when the participants in the study read the text for the

SUMONLY task, their performance in the formation of macrostructure of the texts

was statistically higher (p=.002).

The results indicate that reading a text to complete an MC task does not ensure the formation of the macrostructure of the text and that comprehending the text as a coherent whole is not a byproduct of this test technique. Considering the frequent use of MC tests in the assessment of a variety of constructs of reading ability, the question is whether or not an MC assessment tool serves a valid evaluation technique that is able to evaluate the construct being tested.

Validity of a test is maintained by the fact that it measures what it intends to measure. That is, the validity of a reading test can be evaluated through its ability to reflect whether a test taker is competent in certain constructs of reading. In Chapter 2, it was detailed that reading in an academic context requires that readers construct both a text model of comprehension and a situation model of interpretation (Kintsch, 1998; Grabe, 2009). More specifically, an important skill in an academic context is to be able to "read and understand a text in its entirety with the purpose of learning from it". The skill is associated with high-level reading processes where "reading at the whole-text level" (Khalifa & Weir, 2009) and eventually the formation of the macrostructure of the text is realized. The formation of the macrostructure of a text requires the mastery of global comprehension skills and global comprehension is associated with understanding explicit information in the text and extracting main ideas and making connections between those to eventually integrate and synthesize information (Bax, 2013; Weir & Bax, 2012). Thus, regarding the formation of macrostructure – understanding a lengthy text - as a construct of academic reading ability, a test that intends to measure test takers' academic reading skills should be designed to assess to what extent test takers are competent in the abilities mentioned above.

Given that the current study provided evidence towards the inefficacy of the MC task to direct test takers towards attaining comprehension at textual level, how can decisions concerning their future reading ability in an academic context be made through a test that falls short of assessing the skills necessary for it? If a test is defective in operationalizing the required constructs through which a test taker can reveal the mastery of a skill, it is irrational to expect the test taker to be able to reveal their mastery through that test. In this respect, assessing a test taker's ability to comprehend a text in its entirety through an MC test will not provide construct-related evidence for validity.

The reason why the MC task is regarded as an ineffective test technique to assess macrostructures is that the participants in this study could attain whole-text level comprehension in SUMONLY task but they failed to do so after finishing with the MC task. In this sense, MC tasks can be regarded as being inhibitive in what the participants would otherwise have gained upon reading. This places doubts on the task, not on the participant, concerning the variation in performance and raises questions about the effectiveness of the MC task in assessing test takers' textual level comprehension through the completion of the questions.

What an MC task reveals considering a test taker's reading ability falls within the limits of the questions that appear in MC tests. What this suggests is that there is a limit to what an MC test can offer in terms of the assessment of reading ability and macrostructure formation is beyond its scope with the usual distribution of the questions present in many EFL reading tests. Test takers usually act in the way that is shaped by the requirements of the reading tasks. Reminded by Bachman's method facets (1990) indicating that there are five categories that affect test performance,

two of the categories being "the nature of the expected response" (what the reading task requires the test taker to do) and "the interaction between input and response" (what the test taker does to accomplish the task), we can expect MC test method facets will shape the reading processes of test takers. When the MC task is examined in terms of the method facets, it is fair to say that when test takers are given an MC task, they tend to place the text in a secondary position as they struggle to find an answer for the questions, prioritizing the questions. As Rupp et al. (2006) put it, test takers approach MC tasks as a problem solving task, not a comprehension task. Although the goal should be comprehension, answering the questions correctly replace that goal. This means that it is the task that determines what test takers will end up with upon reading. When an MC test is used to assess academic reading skills, whether the correct answering of the questions lead to the understanding the whole text in remains questionable. In fact, what test takers end up with upon reading a test text may be far from textual level comprehension and there remains a question whether the test measures what it intends to measure: a construct of academic reading; whole text comprehension. Rupp et al. (2006) state that the nature of MC tasks is different from that of the ones in non-testing contexts and that readers go through different processes while they are dealing with such tasks and thus pointing at validity issues regarding MC tasks.

Validity of an assessment task means that the task is a good indicator of what a test taker can do in a real life context. In this sense, a task should, as much as possible, duplicate and correspond to what test takers would do in a real life context. The fact that the nature of MC tasks and the processes carried out for the completion of such tasks are different from the ones in non-testing contexts places further doubts on the validity of the task.

To conclude, we should think over the use of MC tasks for assessment purposes because they are found to be insufficient in directing test takers to read to understand the text and to learn from a text and that they do not help the activation of reading processes or operations test takers will need in order to understand the texts they encounter in real life contexts.

Therefore, it is fair to say that it is not thoroughly possible to assess a test taker's ability to comprehend a text in its entirety with an MC test. As mentioned earlier, the participants in this study were less capable of forming the macrostructure of the text while they read for MC task than they did for SUMONLY task. What this means is that they were not as successful in activating global comprehension skills, extracting main ideas and making connections to integrate and synthesize information. This being the case, MC tasks do not serve as valid assessment techniques to make accurate interpretations about a test taker's academic reading skills.

The findings of the current study indicate that whole text level comprehension is likely to take place when test takers read for the SUMONLY task. That was an expected finding because the selection of SUMONLY task was not made in an arbitrary manner just to choose a counter or an additional task to the MC task. By including a SUMONLY task in the study, we attempted to observe what the participants were able to do in a situation when they had to read and understand a whole text. SUMONLY task design was selected because summaries, as Taylor puts forth (2013) "acknowledge, probably more than most other reading tasks, a view of text comprehension as the construction of a mental representation of the whole text and they therefore offer an appropriate format for assessing this" (p.56).

To be able to assess whether MC tasks were defective in facilitating the attainment of textual level comprehension, a summarization task then was ideal to make the comparison because it is similar to a real life reading activity and it can enable test takers to form macrostructures of the texts.

5.3 Other findings in relation to the scores the participants obtained from the tasks

We think that the following findings obtained in the study also require special mention.

5.3.1 The participants' performance in the tasks

In Chapter 2, it was put forward by several studies that test takers could get relatively higher scores when their reading ability was tested through multiple choice tests (Shohamy, 1984, Wolf, 1991, In'nami & Koizumi, 2009). In the present study, participants performed best in the SUMONLY task although the difference in mean scores of SUMONLY (M=65.57; SD=21.13) and MC (M=63.35; SD=19.71) tasks was minimal.

The fact that the participants in the study scored best in the SUMONLY task needs to be highlighted because summarization as a skill is rarely cultivated in formal education (Brown & Day, 1983) despite many similarities the skill and real life reading have in common. Accordingly, during the sessions with the participants, they frequently stated that summarization was not a familiar and usual activity for them as a reading task. Although the participants encounter MC tasks far more frequently in education and testing settings, they managed to perform better in the SUMONLY task because what they needed to do to summarize the text was quite proximate to a real life reading activity.

The result is, thus, instrumental in explaining what test takers are able to do when the only task is reading and understanding the text to be able to retell it. With the absence of multiple choice questions to answer, test takers can perform the reading activity as naturally as they do in a real life context, without being drifted away by any impediments an MC task may pose.

5.3.2 Extreme cases

While interpreting the analyses of the scores the participants got from the tasks in the current study (see Appendix F for the full list of scores), we noticed that there were a few extreme cases in MC and MCSUM scores (P3, P5, P9 and P10 scored 87.5 in the MC task but their MCSUM scores were only 28.5).

It is clear that although some participants could get a high score in the MC task by answering correctly three quarters of the total questions in the MC task, they summarized the texts quite poorly, displaying the extraction of only about a quarter of main ideas of the texts. In this sense, the MC task seems to be inflating the scores for these participants and leading to incorrect decisions to be made depending on the scores. It is, therefore, assumed that the scores the participants got from the MC task alone would lead to incorrect interpretations and decisions concerning what the participants will be able to do with a reading activity in a real life context.

The extreme difference between the scores of the participants mentioned above may also mean that information intended to be covered in the MC test did not rely on the main ideas to a large extent. Otherwise, the participants' scores in the summaries would have been closer to the scores they got from the MC task.

It should be noted here that it is possible to tap a variety of reading constructs in an MC test with questions that require skills such as literal inferencing, finding references or guessing vocabulary; most of which can be answered through a search reading and/or a local reading activity. What that means is that MC tests that intend to tap several reading sub-skills may deviate from the aim of evaluating global reading ability and the extent to which textual level comprehension can be attained while focusing on much lower sub-skills.

5.4 Research question 2

Research question 2: How do test takers' reading styles and preferences differ according to the tasks (MC and SUMONLY) they are assigned?

Several studies in literature examined reading behaviors of test takers while they are reading and accomplishing tasks (Cohen, 1984, Alderson et al., 1985). Eliciting reading behaviour or operations that test takers carry out during a reading activity is essential to find out whether reading activities operationalize the cognitive processes that they intend to. Only when the observed operations match the ones that are intended to be activated in a test, claims towards validity can be substantiated.

The purpose of RQ2 was to identify the types and frequency of reading operations employed by the participants in the study while they dealt with an MC and a SUMONLY task. By doing so, we aimed at examining the reading operations that are instrumental in the participants' performances during reading for the two tasks and whether the procedures they chose to follow contributed to the formation of the macrostructure of texts.

63

When the responses the participants provided for the verbal protocols were examined, a total of 11 reading operations were put together to form the list of the operations to describe reading processes of the participants (see Table 6). Of the 11 operations in the list, eight reading operations for the MC task (Table 13) and eight for the SUMONLY task (Table 14) appeared to be operationalized by the participants.

5.4.1 Reading operations executed in the MC task

The body of reading operations that the participants stated they went through while completing the MC task point at the characteristics of an MC task and how test takers manage the activity of reading knowing that what they need to do after reading is answering the multiple choice questions.

For the MC task, 93.7% of the participants reported that they "followed a question-to-text sequence, matching the keywords in the text and the questions, (RO3)". That finding confirms the study by Cerdan et al. (2009) who observed their participants following the same process while dealing with MC tasks. Only 21.8% of the participants stated they "read the text carefully before attempting the task (RO1)". That is quite expectable when the nature of MC tasks is taken into consideration; test takers regard MC tasks as a "problem-solving activity", where the problem is the question and the text is only a means to answer it. That perspective in reading puts the text in a secondary position, prioritizing the question(s). The test taker' goal in reading is, then, to find an answer that is acceptable, rather than to understand the text.

Following a question-to text sequence in test taking also means that reading the whole text may not be a requirement, which is supported by the data that show only 37.5% of the participants stated they "have read the whole text from the beginning to the end carefully (RO6)". That is a critical observation to make in terms of reading activities. When the whole text is not read, as proven by the remaining 62.5% of participants who did not mention reading the text in full, considerations towards a representative sample of comprehension abilities cannot be made. Reading for answering the questions, that is the logic of MC design, seems to contribute to finding the answers to the questions as test takers may devote less effort for textual comprehension. If a reading task enables test takers to complete the task by answering the questions without reading the whole text, the task is more like a jigsaw activity where test takers find out which information in a text fits the question. Attaining comprehension at textual level by macrostructure formation requires a careful, high-level reading process where readers are able to use their ability to integrate information and draw conclusions (Pressley, 2002). It was reported by almost a quarter of the participants that they "read expeditiously to find a relevant part that is thought to include the answer (RO4)" and 28.1% of them expressed that they "read carefully only the selected part(s) of the text that is thought to be relevant to the questions (RO5)". Careful reading is critical to reach textual understanding; however, it should also apply to the whole text. Thus, when a careful and a linear reading style is not adopted, and most importantly, when reading activity takes place in a segmental or partial manner, formation of macrostructure is not likely to take place because the processes necessary for it cannot be substantiated during such a reading activity as was confirmed by MCSUM test results.

A local and segmental style of reading, whether it is done in a careful or expeditious manner, breaks up the unity in a reading material and leaves test takers with bits and pieces of information. Such distortion in the integrity and connectedness of the parts and paragraphs of a text as a whole makes it impossible for the test taker to have all recombined upon finishing with the task. In this study, only one out of 32 participants mentioned s/he "tries to understand how the text is organized, how the ideas and details connect each other (RO9)", indicating that almost none of the participants thought understanding text organization was important for MC task completion.

As the rationale of reading to find answers does not correspond to that of macrostructure formation, which is the assimilation of the information into the reader's existing knowledge (Kintsch, 1994), reading activity outcomes derived from the tasks are likely to be incongruent.

To conclude, for MC task completion, the participants mostly made use of the question-driven or strategic operations during reading. The reading operations that were uniquely operationalized in the MC task, but were not mentioned of in the SUMONLY task were RO3 (follows a question-to-text sequence), RO4 (reads expeditiously to find a relevant part that is thought to include the answer) and RO5 (reads carefully only the selected part of the text that is thought to be relevant to the question). As we said before, text level understanding might be hindered through task-specific strategic question answering in a multiple choice task whereas in linear, careful reading to summarize a text, as there are no interfering processes, text level understanding is more possible. Therefore, it is possible to underline, once more that task specific MC reading operations do not contribute to a deeper understanding of a text; on the contrary, they may even be hindering it.

5.4.2 Reading operations activated during the SUMONLY task

While reading for the SUMONLY task, 96.8% of the participants stated they "read the text carefully first, before attempting the task, (RO1)" and as a result, they "have read the whole text carefully, (RO6)". That means one important main objective what a reading assessment should have is accomplished: making test takers read the text in full. As naturally expected, this overwhelming percentage points at a careful, linear reading style that is done when there are no questions that follow the text. Such a reading activity can be considered as a natural, authentic reading that involves only the interaction of the reader and the text. In this respect, it resembles the reading skills readers use in real life contexts, on which tests of various kinds attempt to make inferences.

As Grabe (2009) pointed out, text model comprehension of the text is developed by extracting the main ideas and forming a coherent whole of them and then creating the situation model of interpretation by internalizing the text to retell it in readers' own words. 37.5% of the participants reported that they "read to get the main ideas and remember them, (RO10)" and 21.8% of them said they "paid further attention to the introduction and conclusion paragraphs as they are thought to include the main idea of the whole text, (RO11)" when they are dealing with SUMONLY task. During summarization, high level processes come into play (Khalifa & Weir, 2009), directing the reading activity so that it follows a global and holistic manner. 31.2% of the participants asserted that they "read to make connections between paragraphs or parts, (RO8)" and 15.6% of them said they "try to understand how the text is organized and how the ideas and details connect each other, (RO9)" to understand the text. These exemplify in their wording such a reading style was in use.

To underline once again, the reading operations that were particular to the SUMONLY task were RO9 (tries to understand how the text is organized and how the ideas and details connect each other), RO10 (reads to get the main ideas and remember them) and RO11 (pays further attention to the introduction and conclusion paragraphs as they are thought to include the main idea of the whole text). It is clear that they are genuinely text-driven operations a test taker can make use of to attain comprehension at text level.

When those operations do not exist in the reading process as in the case of MC tasks in this study, the achievement of the formation of macrostructure of a text cannot be expected. That could explain why the participants dealing with the SUMONLY task did better in the formation of macrostructure after reading in comparison to MC task.

5.5 Conclusion

The investigation of RQ1 put forward that MC tasks act as an impediment to the attainment of whole text level comprehension and the analysis of RQ2 showed us that they direct test takers to adopt an MC-specific, strategic reading process. This study raised an important point on the invalidity of reading assessment tests that are formed of MC questions as such tests do not seem to give us information about whether a test taker can understand an extended text, for example a page-long text, but they can inform us only on the ability of searching for information and reading locally. This can have a negative impact both on the accuracy of our decisions based on the test scores and on teaching where high-stakes tests dominate language learning. As such, widespread use of the MC technique should be questioned once again.

CHAPTER 6

CONCLUSION

6.1 Introduction

The present study was conducted to investigate whether test takers would be able to attain whole text level comprehension upon the completion of two tasks: MC and SUMONLY. We attempted to assess their understanding of the texts at whole text level by comparing the scores they received from the MCSUM (the summary of the text they read to complete the MC task) and the SUMONLY (the baseline task). In addition to the performance they had in the formation of macrostructures of texts, how they read the texts and whether they opted for different reading operations while reading for the two tasks were observed.

Analyses carried out from the data gathered show that the performance the participants displayed in terms of macrostructure formation of texts after completing both tasks differed significantly and that they were less successful in comprehending the text in its entirety although they could get second best result from the MC test (Table 8).

Regarding the reading operations that the participants executed, there emerged quite diverse selection of operations to be used. For the MC task, the participants acted more strategically and implemented MC-specific operations such as following a question-to-text sequence and searching and locating information that could lead to answering the questions whereas during reading for the SUMONLY task, the participants followed a more real-life like reading process; reading to understand and elicit main ideas.

There is a connection between the two findings of the study. The fact that the participants could not perform as well in the MCSUM task can be explained through the nature of the MC task, which did not direct the participants towards utilizing reading operations that would assist them in macrostructure formation. However, in SUMONLY task, the participants were able to reach whole text level comprehension because they picked the reading operations necessary to process a text fully, above average and better than MCSUM task.

6.2 Implications for teaching and testing of the reading skill

The results of the study have certain implications on the teaching and testing of the reading skill:

In academic contexts, what readers need to be equipped with in terms of the reading ability they possess has been discussed in Chapter 2. Taking the reading construct in academic contexts into account, classroom instruction and teaching programs preparing learners for academic life should incorporate practices that cultivate and enhance the required academic reading skills. The foci of classroom practices and materials should be designed to teach learners how to read texts to perform well in different reading types and process a text fully to extract complete meanings from (Weir et al., 2009) and create a text and a situation model (Kintsch, 1998) by reading at the whole text level (Khalifa & Weir, 2009).

The current study provided evidence for the inefficacy of MC questions towards comprehension at textual level. This being the case, the frequent use of MC tasks in educational settings should be questioned and educators should be aware of the shortcomings of the task in helping learners towards the mastery of reading skills.

Due to the practicality of MC tasks in terms of preparation and administration, they will definitely take place in educational settings but attention should be paid whether the purposes of educational practices can be fulfilled though the use of MC tasks. There are, definitely, several reading skills and subskills to be fostered through MC tasks. Reading skills is an umbrella term that covers many competencies a learner needs to develop to be able to cope with the written materials that s/he encounters in real life. MC tasks, for instance, can be regarded as effective tools to teach and enhance the reading skills that require the mastery of a local, segmental reading ability, the ability to scan or skim for information or even to form a microstructure at paragraph level. To teach textual level comprehension skills, however, other tasks or question techniques that can contribute to that specific purpose should be utilized. That is, teaching practices should be valid in that they serve for the development of the required result, whatever it is that educators attempt to cultivate in learners. As accurate interpretations about a test taker's ability in an aspect of language cannot be made depending on a test that is speculative in measuring what is intended to be measured, it is not realistic to expect desired outcomes from teaching programs that include practices that are questionable in terms of teaching what is intended to be taught.

As proven to be instrumental in the attainment of textual level comprehension in the current study, summaries, both oral or written, should be utilized more for teaching purposes. Summaries are good indicators of the formation of macrostructures and the macrostructure of the written material reflects reading comprehension and ability more effectively (Kintsch & Kintsch, 2005).

In addition, as the current study found evidence towards the importance of the execution of the necessary reading operations in order to arrive at whole text level understanding, reading strategies incorporating such reading operations should be highlighted and taught.

For the assessment procedures, we need to make sure decisions concerning the technique to be used in tests are not made at the expense of validity. Therefore, it is obvious that tests designed for academic purposes should be assessing test takers' reading ability in terms of the constructs mentioned for real life academic reading skills. However, as the findings of the current study revealed, one of the most common test techniques, the multiple choice format, seems to be deficient in providing insights about how well test takers will perform in tasks that require the operation of the constructs of academic reading. In tests that assess academic reading skills, techniques, question types or items that target test takers' ability to reflect the constructs, i.e. extracting complete meaning, forming the macrostructure, attaining at textual level comprehension should be included. Otherwise, we end up assessing a reading skill that includes sub-skills of reading at more local levels which are less relevant to academic settings where extended texts need to be processed and understood. This is a serious validity issue.

One should also keep in mind that reading comprehension is assessed in many research studies in the area of Second Language Acquisition and others as an important variable. If such studies use MC format tasks only, they are also risking the validity of the inferences that could be made from those studies.

6.3 Limitations of the study and suggestions for further research

The major limitation of the study was the small sample size. Because of the limited amount of data, the conclusions drawn in this study are tentative and preliminary.

Of the 32 participants in the study, 11 of them were already the students of the researcher. Therefore, the study was not a double-blinded one. However, when educational and testing settings are taken into consideration, that is often the case. To alleviate the possible drawbacks of this, inter-rater agreement for the coding of reading operations and intra-rater agreement for the checking of summary scores were called for.

During verbal protocols, what the participants expressed concerning the processes they followed when they were completing the tasks were operationalized into reading operation categories in this study also through comparing them to the reading operation schemes in other studies. More studies investigating test taker reading styles and preferences may provide additional operations that were not included in the current study. Probing into how test takers read along with eye-tracking data may provide more insights on how reading operates in their minds.

Given that substantial evidence towards the deficiency of MC tasks in assessing comprehension at textual level has been presented in the current study, there appears a need for future research that focuses on creating innovative test techniques to assess comprehension at textual level. Summaries used in the study were found to be contributory in whole text level comprehension but they are not practical in terms of assessment purposes. Therefore, assessment techniques that are goal-driven and practical to use but also cognitively valid should be developed in the future.

APPENDIX A

INFORMED CONSENT FORM

KATILIMCI BİLGİ ve ONAM FORMU

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi

Araştırmanın adı: Assessment of textual level EFL reading comprehension: a
cognitive and contextual investigation. / Okuma becerilerinin yabancı dil olarak
İngilizce bağlamında metin seviyesinde bilişsel ve bağlamsal açıdan
değerlendirilmesi

Proje Yürütücüsü: Aylin Ünaldı (Boğaziçi Üniversitesi, Eğitim Fakültesi)

Kurumsal adres: Boğaziçi Üniversitesi, Eğitim Fakültesi, Yabancı Diller eğitimi
bölümü, 34342 Bebek İstanbul Telefon: 0212 3594609 e-posta:

aunaldi@boun.edu.tr

Araştırmacının adı: Burcu Kayarkaya

E-mail adresi: burcudurak@hotmail.com

Telefonu: 505 274 39 80

Proje konusu: Okuma becerileri yabancı dil yeterliği/seviyesi belirleme sınavlarında
ölçülen dört beceriden biridir. En basit anlamıyla bu tür sınavlar, kişinin "okuduğunu
anlama" becerisine ne ölçüde hakim olduğunu tespit etmek amacını güder. Bu amaç
doğrultusunda, belli kriterler kapsamında geliştirilen okuma metinleri, farklı soru
tipleri kullanılarak yabancı dil becerisi ölçen sınavların okuma bölümlerinin içeriğini
oluşturur. Bu araştırmanın amacı bu metinler, hedef kitleleri tarafından ne şekilde
okunduklarını ortaya çıkarmaktır. Metni okuduktan sonra ne yapması gerektiği
yönergelerle belirtilen okurların aynı bilişsel süreçlerden geçerek okuma eylemini
tamamlayıp tamamlamadıklarını tespit etmek, araştırmacının temel sorularını
yanıtlama konusunda yardımcı olacaktır. Mevcut araştırmada, metinleri takiben
verilecek iki farklı soru tipi (çoktan seçmeli ve sözlü özet) kapsamında proje
katılımcılarının bilişsel süreçlerinde farklılık gözlenip gözlenmediğinin ortaya
çıkarılması hedeflenmektedir.

Onam: Bu çalışmanın sonucunda, çalışmada kullanılan test formatları dahilinde,
okuma parçalarının sonrasında gelen soru tiplerine bağlı olarak katılımcıların okuma
biçimlerinin ve bilişsel düzeydeki faaliyetlerinin değişip değişmediği ve herhangi bir
soru tipinde katılımcıların okuma becerilerinin yansıtılmasında bir üstünlük gözlenip
gözlenmediği hakkında bilgi edinmeyi hedefliyoruz.

Araştırmaya katılmayı kabul ettiğiniz takdirde, Yıldız Teknik Üniversitesi
Davutpaşa Kampüsünde birlikte belirleyeceğimiz bir sınav salonunda, yaklaşık olarak

2 saat sürecek bir çalışmaya bireysel olarak katılacaksınız. Bu çalışma kapsamında, biri çoktan seçmeli sorulardan oluşan, biri ise okuma sonrası sözlü olarak özetlemeniz gereken 2 okuma sınavı çözeceksiniz. Bu sınavları çözerken mümkün olduğu kadar konuşarak aklınızdan geçenleri, izlenimlerinizi ve hissettiklerinizi ifade ederek sesli bir biçimde düşünmeniz istenecektir. Araştırmacı size bu konuda yardımcı olmak için sizi yönlendirecek sorular soracak ve veriyi daha sonra analiz edebilmek için tüm oturumu kayıt altına alacaktır.

Bu çalışma boyunca, isminiz ve kayıtlar tamamen gizli tutulacaktır ve çalışmanın herhangi bir aşamasında isminiz kullanılmayacaktır.

Çalışmaya katılmanız tamamen isteğe bağlıdır. Bu çalışmaya katılıp katılmamanız ders notlarınızı hiçbir şekilde – olumlu ya da olumsuz – etkilemeyecektir. Sizden ücret talep etmiyoruz ancak ayıracağınız vakit, bilimsel bir çalışmanın gerçekleştirilmesine katkıda bulunacaktır. Ayrıca çalışmaya katılmanızın karşılığında size "Boğaziçi Üniversitesi" yazılı mug ya da YTU kampüsteki kafeteryada kullanılmak üzere bir yemek fişi hediye edilecektir.

Sizden alınan örnek ileride başka çalışmalar için de kullanılabilir. Katıldığınız taktirde çalışmanın herhangi bir aşamasında herhangi bir sebep göstermeden onayınızı çekmek hakkına da sahipsiniz. Yapmak istediğimiz araştırmanın öngörülen hiçbir olumsuz etkisi ya da ayıracağınız 2 saat dışında size hiçbir yükü olmayacaktır. Çalışma sırasında yapılan analiz sonuçları isterseniz kişisel olarak sizinle paylaşılacaktır.

Araştırma sırasında ve sonucunda ortaya çıkan bilgiler bizi farklı amaçlar için (farklı soru tiplerini cevaplamak için) okunan okuma parçalarının, bilişsel düzeyde okurun (katılımcıların) farklı süreçlerden geçip geçmediğini ve fark saptandığı taktirde, herhangi bir soru tipinin diğerlerine göre okuma becerilerini yansıtması açısından daha etkili olup olmadığını gösterecek.

Bu bilgiler, size de kişisel olarak ileri düzeyde okuma becerinizi yansıtmada ve geçtiğiniz bilişsel süreçlerin farkında olup daha etkin bir okur olmak için gereken adımları atmanızda yardımcı olabilir.

Bu formu imzalamadan önce, çalışmayla ilgili sorularınız varsa lütfen sorun. Daha sonra sorunuz olursa, araştırmacı Burcu Kayarkaya'ya (Telefon: 0505 274 3980; e-mail: burcudurak@hotmail.com) ya da proje yürütücüsü Aylin Ünaldı'ya (e-mail: aunaldi@boun.edu.tr) sorabilirsiniz. Araştırmayla ilgili haklarınız konusunda Boğaziçi Üniversitesi İnsan Araştırmaları Etik Alt Kurulu (INAREK) veya INAREK/SBB Etik Alt Kurulu kurullarına da danışabilirsiniz.

Bana anlatılanları ve yukarıda yazılanları anladım. Bu formun bir örneğini aldım. Çalışmaya katılmayı kabul ediyorum.

Katılımcı Adı-Soyadı:…………………………………..
İmzası: ……………………………………………
Tarih (gün/ay/yıl):........./.........../..............

Text A-MC

Text A-SUMONLY is the version with the same text without the questions. Time allotted for the SUMONLY task: 10 mins.

Instruction provided by the researcher to the participant:
Read the text and answer the questions below. You have 20 mins. for this task.

WHICH HAND DID THEY USE?

A   We all know that many more people are right-handed than left-handed. Can one trace this same pattern far back in prehistory? (1A) Much of the evidence about right-hand versus left-hand dominance comes from stencils and prints found in rock shelters in Australia and elsewhere, and in many Ice Age caves in France, Spain, and Tasmania. (1B) When a left hand has been printed on a surface, this implies that the artist was right-handed, and vice versa. (1C) Even though the paint was often sprayed on by mouth, one can assume that the dominant hand assisted in the operation. One also has to make the assumption that hands were printed palm downward—a left hand printed palm upward might of course look as if it were a right hand. (1D) Of 158 stencils in the French cave of Gargas, 136 have been identified as left, and only 22 as right; right-handedness was therefore heavily predominant.

B   Cave art furnishes other types of evidence of this phenomenon. Most engravings, for example, are best lit from the left, as it is appropriate for the work of right-handed artists, who generally prefer to have the light source on the left so that the shadow of their hand does not fall on the tip of the engraving tool or brush. In the cases where an Ice Age figure is depicted holding something, it is mostly, though not always, in the right hand.

C   Clues to right-handedness can also be found by other methods. Right-handers tend to have longer, stronger, and more muscular bones on the right side. Marcellin Boule, a French paleontologist, noted after an excavation he directed that the La Chapelle-aux-Saints Neanderthal skeleton had a right upper arm bone that was noticeably stronger than the left. Similar observations have been made on other Neanderthal skeletons such as La Ferrassie I and Neanderthal itself.

D   Fractures and other cut marks are another source of evidence. Right-handed soldiers tend to be wounded on the left. The skeleton of a 40- or 50-year-old Nabatean warrior, buried some 2,000 years ago in the Negev Desert, Israel, had multiple healed fractures to the skull, the left arm, and the ribs.

E   Tools themselves can be revealing, too. Long-handed Neolithic spoons made of wood preserved in Alpine villages dating to 3000 B.C. have survived; the signs of rubbing on their left side indicate that their users were right-handed. The late Ice Age rope found in

the French cave of Lascaux consists of fibers spiraling to the right, and was therefore woven by a right-hander.

F   Occasionally one can determine whether stone tools were used in the right hand or the left, and it is even possible to assess how far back this feature can be traced. In stone toolmaking experiments, Nick Toth, a right-hander, tried to act as if he had been the one who produced the tools so that he could see whether the real producer was a right or a left-hander. He held the core (the stone that would become the tool) in his left hand and the hammer stone (the stone used to produce the tool) in his right. Toth's experiment on the tool produced a result of 56 percent right-oriented strikes while creating the tool. Toth has applied the same criteria to the similarly made stone equipment from a number of early sites at Koobi Fora, Kenya. At seven sites he found that 57 percent of the flakes were right-oriented, and 43 percent left, a pattern almost identical to that produced today.

G   About 90 percent of modern humans are right-handed: we are the only living creature with a varied use of one hand. The part of the brain responsible for fine control and movement is located in the left cerebral hemisphere, and the findings above suggest that the human brain was already asymmetrical in its structure and function not long after 2 million years ago. Among Neanderthalers of 70,000-35,000 years ago, Marcellin Boule noted that the La Chapelle-aux-Saints individual had a left hemisphere slightly bigger than the right, and the same was found for brains of specimens from Neanderthal, Gibraltar, and La Quina.

1 All of the following are mentioned in paragraphs A and B as evidence of right-handedness in art and artists EXCEPT

A the ideal source of lighting for most engravings
B the fact that a left hand stenciled palm upward might look like a right arm
C the prevalence of outlines of left hands
D figures in prehistoric art holding objects with the right hand

2 Look at the four parenthesis ( ) that show where the following sentence could be added to paragraph A.

The stencils of hands found in these shelters and caves allow us to draw conclusions about which hand was dominant.

A (1A)   B (1B)   C (1C)   D (1D)

3 According to paragraph C, the La Chapelle-aux-Saints Neanderthal skeleton can be identified as right-handed because

A other Neanderthal skeletons found in the area nearby are also right-handed
B the right arm bone is stronger than the left arm bone
C it is similar to skeletons of La Ferrasie I and Neanderthal
D the right side of the skeleton shows less evidence of fractures

4 Which of the following statements about fractures and cut marks can be inferred from paragraph D?

A Fractures and cut marks caused by right-handed soldiers tend to occur on the right side of the injured party's body.

B The right arm sustains more injuries because, as the dominant arm, it is used more actively.
C In most people, the left side of the body is more vulnerable to injury since it is not defended effectively by the dominant arm.
D Fractures and cut marks on fossil humans probably occurred after death.

5 In paragraph E, why does the author mention the Ice Age rope found in the French cave of Lascaux?

A as an example of an item on which the marks of wear imply that it was used by a right-handed person
B because tressing is an activity that is easier for a right-handed person than for a left-handed person
C because the cave of Lascaux is the site where researchers have found several prehistoric tools made for right-handed people
D as an example of an item whose construction shows that it was made by a right-handed person

6 The word "that" in paragraph F refers to

A criteria   B similarly made stone equipment
C a number of early sites   D the percent of right and left handers

7 Examine the sentence in paragraph G below:

About 90 percent of modern humans are right-handed: we are the only living creature with a preferential use of one hand.
Which sentence below most clearly expresses important information in the sentence above? Incorrect choices change the meaning or leave out important information.

A Most modern humans are right-handed; and that property is imprinted in their genes as well as the animals'.
B About 90 percent of humans today preferentially use their right hand although otherwise is possible.
C Almost all modern humans are right-handed; and they differ from other animals in their variedness.
D Only 10 percent of humans use their left hand in the modern world, but it is not the case for the animals.

8 The word "specimens" in paragraph G could be best replaced by

A remains
B researchers
C right-handers
D left-handers

Please indicate on the five-point scale below by circling the relevant number:

How familiar are you with the topic of the text or the information you have just read?
1   2   3   4   5

(1: I had never heard about it before. / 5: I was very familiar with it.)

Text difficulty – How difficult is the text for you?
1   2   3   4   5

(1: It is very easy. / 5: It is very difficult)

Text B-MC

Text B-SUMONLY is the version with the same text without the questions. Time allotted for the SUMONLY task: 10 mins

Instruction provided by the researcher to the participant:
Read the text and answer the questions below. You have 20 mins. for this task.

TRAINING THE BRAIN

A   People who can accomplish unbelievable mnemonic feats, such as memorizing thousands of random digits in under an hour, claim they have normal brains and that they do not have photographic memories: a gift that some people are born with that enables them to remember anything and everything. Some of these memory superstars compete annually in Olympic-like World Memory Championships. What these athletes do utilize are techniques that anyone can incorporate into everyday life to train one's memory. In addition to using techniques, these competitors undergo serious training and practice.

B   The World Memory Championships begin with the competitors sitting at the table with two shuffled decks of cards. Each person will have exactly five minutes to memorize the order of both decks. These mental athletes, or MAs in short, can memorize the first and last names of dozens of strangers or any poem handed them in only a few minutes. Ed Cooke, a 24-year-old MA from England, explains that MAs see themselves as participants in an amateur research program attempting to rescue the long-lost art of memory training. In the not so distant past, Cooke contends, culture depended on individual memories. Almost all of Cooke's mnemonic techniques were invented in ancient Greece. These techniques existed not to recall useless information, such as playing cards, but to carve into the brain foundational texts and ideas.

C   A study in the journal Nature examined eight of the people who finished near the top of the World Memory Championships. The scientists examined whether these contestants' brains were fundamentally different from everyone else's or whether these people were simply making better use of memorizing abilities that we all possess. The researchers put the MAs and control subjects into brain scanners and had them memorize numbers, photographs of people, and snowflakes. What they found was astonishing. The brains of the MAs and those of the control subjects were anatomically indistinguishable. On every test of mental ability, the MAs scored in the normal range. One surprising difference between the two groups surfaced; when the researchers examined what part of the brain was utilized during a memory activity, they found the MAs relied more heavily on regions in the brain involved in spatial memory.

D   MAs offer a simple explanation. (5A) Anything can be imprinted upon our memories and kept in good order, simply by constructing a building in the imagination and filling it with images of what needs to be recalled. Dating back to the fifth century, this building is called a Memory Palace. (5B) Even as late as the fourteenth century, when there were perhaps only a dozen copies of any text, scholars needed to remember what was read or told to them. (5C) Reading to remember requires a very different technique than speed reading. If something is going to be made memorable, it has to be repeated. Until relatively recently, people read only a few books intensively over and over again, usually out loud in groups. (5D) Today we read extensively, usually only once and without sustained focus.

E   What distinguishes the great mnemonist is the ability to create lavish images on the spur of the moment, to paint a picture in one's mind so unlike any other it cannot be forgotten and to do it quickly. Using memory palaces, contestants create memorized images. Take a deck of cards, for example, and recombine the pictures to form unforgettable scenes such as routes through a town or signs of the zodiac. One competitor used his own body parts to help him memorize the entire 57,000-word Oxford English-Chinese dictionary.

F   Any novice who wishes to train the mind needs first to stockpile palaces. By visiting the homes of old friends, taking walks through museums, or compiling a collection of famous artists, one can build new, fantastical structures in the imagination. Then carve each building up into cubbyholes for memories. One mnemonist associates every card in deck with a different celebrity performing a strange act. Another puts every card into a different exhibit at her favorite museum. In a short period of time, one will notice improvement with remembering license plate numbers or shopping lists. In order to keep the skill sharp, MAs deliberately empty their palaces after competitions, so they can reuse them again and again and recommend that novices do the same.

1 The word "undergo" in paragraph A could be best replaced by

A participate in   B contemplate
C transport   D travel to

2 According to paragraph B

A mental athletes try to revive the forgotten practice of memory training
B mental athletes are trained to memorize poems or decks of cards
C ancient Greeks used mnemonic techniques to play cards
D ancient Greeks memorized first and last names of strangers

3 The word "those" in paragraph C refers to

A photographs of people   B snowflakes
C brains   D MAs

4 According to paragraph C, it is NOT TRUE that mental athletes

A score in the normal range of mental ability tests
B have brains that are anatomically different from everyone else's
C depend more on areas of the brain that control spatial memory
D are compared to a control group to examine their memorizing abilities

5 Look at the four parenthesis ( ) that show where the following sentence could be added to paragraph D.

They were able to recall large amounts of information by storing information in such a structure.

A (5A)   B (5B)   C (5C)   D (5D)

6 Why does the writer mention "speed reading" in paragraph D?

A to discuss a fourth century technique
B to illustrate why people read a few books intensively

C to explain the copies of texts fourteenth century scholars needed to recall
D to contrast the type of reading done nowadays with that of earlier times

7 It can be inferred from paragraph E that

A there are 57,000 words in Chinese language
B there is a variety of unforgettable scenes people can create
C memory palaces can easily be forgotten
D deck of cards can build actual buildings

8 Examine the sentence in paragraph F below:

Any novice who wishes to train the mind needs first to stockpile palaces.

Which sentence below most clearly expresses important information in the sentence above?
Incorrect choices change the meaning or leave out important information.

A Those new to memory training need to create multiple memory palaces.
B Stockpiling memory palaces enables those new to memory competitions to win.
C Training the mind happens when one is new to competitions.
D When one stockpiles in battle, one enters a memory palace.

Please indicate on the five-point scale below by circling the relevant number:

How familiar are you with the topic of the text or the information you have just read?

1   2   3   4   5

(1: I had never heard about it before. / 5: I was very familiar with it.)

Text difficulty – How difficult is the text for you?

1   2   3   4   5

(1: It is very easy. / 5: It is very difficult)

RUBRICS FOR SUMMARIES

Rubric for Text A - Which hand did they use?

1. The writer is interested in whether the number of right-handers exceeded that of left-handers in the past.

2. To achieve this, cave art is examined by researchers (the works of ancient people).

3. Work on excavations is another source of information where evidence towards right or left handedness can be found.

4. Fractures and cut marks are other sources of evidence

5. Examining the tools used by ancient people can also provide insight.

6. (Nick Toth's) experiment emphasizes that the rate of right vs left handedness in the world has not changed throughout ages.

7. Almost all modern humans are right-handed; and they differ from other animals in their changing ability to use either hand.

Rubric for Text B - Training the brain

1. The writer wants to explore why some people have sharper memories.

2. It (photographic memory) is a gift but at the same time it can be achieved through training.

3. (Study shows that) MAs have photographic memories or spatial ability and they have normal brains.

4. MAs compete in World Memory Championships (Olympics).

5. MAs use techniques that date back to ancient Greeks. MAs try to revive the long lost art of memory training.

6. MAs use Memory Palaces and load them with mental representations of what they want to remember. (They are detailed and unique structures, and are as rich as one's imagination allow them to be.)

7. Novices are advised that they have readily available palaces in mind so that they can fill in whenever necessary to remember easily.

APPENDIX D

INTRA-RATER AGREEMENT FOR THE SUMMARY SCORES

Table D1. Intra-rater Agreement for the MCSUM Scores

Crosstabs – Symmetric measures

	value	asymptotic standard error	approximate T	approximate significance
Measure Kappa of agreement	.76	.08	9.406	.00
N of valid cases	32			

Table D2. Intra-rater Agreement for the SUMONLY Scores

Crosstabs – Symmetric measures

	value	asymptotic standard error	approximate T	approximate significance
Measure Kappa of agreement	.73	.09	8.767	.00
N of valid cases	32			

READING OPERATIONS BY RATER- MC TASK

Table E1. Rater 1-Reading Operations for the MC Task

Inter-rater agreement: 81%

RATER 1 READING OP. FOR MC TASK	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1			x			x	x				
PARTICIPANT 2		x	x	x		x					
PARTICIPANT 3	x		x		x	x	x				
PARTICIPANT 4			x	x			x				
PARTICIPANT 5	x					x	x				
PARTICIPANT 6			x		x						
PARTICIPANT 7		x	x			x					
PARTICIPANT 8			x	x		x					
PARTICIPANT 9		x	x								
PARTICIPANT 10	x		x			x					
PARTICIPANT 11	x		x					x			
PARTICIPANT 12			x	x	x						
PARTICIPANT 13			x			x					
PARTICIPANT 14	x		x								
PARTICIPANT 15	x		x								
PARTICIPANT 16			x		x						
PARTICIPANT 17			x		x						
PARTICIPANT 18	x		x		x	x					
PARTICIPANT 19			x				x				
PARTICIPANT 20			x								
PARTICIPANT 21		x	x								
PARTICIPANT 22			x				x				
PARTICIPANT 23			x		x						
PARTICIPANT 24			x								
PARTICIPANT 25	x		x		x	x					
PARTICIPANT 26			x								
PARTICIPANT 27	x			x	x	x					
PARTICIPANT 28		x	x		x						
PARTICIPANT 29		x	x								
PARTICIPANT 30			x				x				
PARTICIPANT 31			x		x		x				
PARTICIPANT 32			x	x	x	x					

Table E2. Rater 2-Reading Operations for the MC Task

RATER 2											
READING OP. FOR MC TASK	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1			x		x	x	x				
PARTICIPANT 2		x	x	x	x	x					
PARTICIPANT 3	x		x			x	x				
PARTICIPANT 4			x	x			x				
PARTICIPANT 5	x					x	x				
PARTICIPANT 6			x		x		x				
PARTICIPANT 7		x	x			x					
PARTICIPANT 8			x	x	x	x					
PARTICIPANT 9		x	x								
PARTICIPANT 10	x		x			x					
PARTICIPANT 11	x		x			x		x			
PARTICIPANT 12			x	x	x						
PARTICIPANT 13			x			x	x				
PARTICIPANT 14			x		x						
PARTICIPANT 15			x			x					
PARTICIPANT 16			x		x						
PARTICIPANT 17			x			x	x				
PARTICIPANT 18	x		x			x	x				
PARTICIPANT 19			x				x				
PARTICIPANT 20			x								
PARTICIPANT 21		x	x								
PARTICIPANT 22			x				x				
PARTICIPANT 23			x		x						
PARTICIPANT 24			x								
PARTICIPANT 25	x		x			x					
PARTICIPANT 26	x		x								
PARTICIPANT 27	x			x		x					
PARTICIPANT 28		x	x		x						
PARTICIPANT 29		x	x								
PARTICIPANT 30			x				x				
PARTICIPANT 31			x			x	x				
PARTICIPANT 32			x	x	x						

APPENDIX F

READING OPERATIONS BY RATER- SUMONLY TASK

Table F1. Rater 1-Reading Operations for the SUMONLY Task

Inter-rater agreement: 73%

RATER 1 READING OP. FOR SUMMARY T.	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1	x					x				x	
PARTICIPANT 2	x					x				x	x
PARTICIPANT 3	x					x	x				
PARTICIPANT 4	x					x		x			x
PARTICIPANT 5	x					x				x	
PARTICIPANT 6	x					x		x			
PARTICIPANT 7	x					x			x		
PARTICIPANT 8	x					x	x			x	x
PARTICIPANT 9	x					x					
PARTICIPANT 10	x					x				x	
PARTICIPANT 11	x					x	x	x	x		x
PARTICIPANT 12	x					x	x				
PARTICIPANT 13	x					x					
PARTICIPANT 14	x					x		x		x	
PARTICIPANT 15	x					x					
PARTICIPANT 16	x					x					
PARTICIPANT 17	x					x				x	
PARTICIPANT 18	x					x	x	x		x	x
PARTICIPANT 19	x					x					
PARTICIPANT 20	x					x	x				
PARTICIPANT 21		x					x		x		x
PARTICIPANT 22	x					x					
PARTICIPANT 23	x					x					
PARTICIPANT 24	x					x		x			
PARTICIPANT 25	x					x	x	x	x		x
PARTICIPANT 26	x					x					
PARTICIPANT 27	x					x		x		x	
PARTICIPANT 28	x					x					
PARTICIPANT 29	x					x				x	
PARTICIPANT 30	x					x	x		x	x	
PARTICIPANT 31	x					x				x	x
PARTICIPANT 32	x					x		x	x		

Table F2. Rater 2-Reading Operations for the SUMONLY Task

RATER 2 READING OP. FOR SUMMARY T.	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1	x					x		x		x	
PARTICIPANT 2	x					x		x		x	x
PARTICIPANT 3	x					x	x			x	
PARTICIPANT 4	x					x			x		x
PARTICIPANT 5	x					x		x		x	
PARTICIPANT 6	x					x					
PARTICIPANT 7	x					x		x		x	
PARTICIPANT 8	x					x	x				x
PARTICIPANT 9	x					x				x	
PARTICIPANT 10	x					x		x			
PARTICIPANT 11	x					x	x		x		x
PARTICIPANT 12	x					x		x			
PARTICIPANT 13	x					x					
PARTICIPANT 14	x					x		x	x	x	
PARTICIPANT 15	x					x	x				
PARTICIPANT 16	x					x		x			
PARTICIPANT 17	x					x				x	
PARTICIPANT 18	x					x			x		x
PARTICIPANT 19	x					x	x				
PARTICIPANT 20	x					x	x				
PARTICIPANT 21		x						x		x	x
PARTICIPANT 22	x					x				x	
PARTICIPANT 23	x					x		x			
PARTICIPANT 24	x					x	x			x	
PARTICIPANT 25	x					x	x	x		x	x
PARTICIPANT 26	x					x					
PARTICIPANT 27	x					x		x		x	
PARTICIPANT 28	x					x					
PARTICIPANT 29	x					x		x		x	
PARTICIPANT 30	x					x	x	x			
PARTICIPANT 31	x					x		x		x	x
PARTICIPANT 32	x					x	x	x		x	

READING OPERATIONS AGREED BY RATERS

Table G1. Reading Operations Agreed by Raters for the MC Task

READING OP. FOR MC TASK	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1			x			x	x				
PARTICIPANT 2		x	x	x		x					
PARTICIPANT 3	x		x			x	x				
PARTICIPANT 4			x	x			x				
PARTICIPANT 5	x					x	x				
PARTICIPANT 6			x		x						
PARTICIPANT 7		x	x			x					
PARTICIPANT 8			x	x	x	x					
PARTICIPANT 9		x	x								
PARTICIPANT 10	x		x			x					
PARTICIPANT 11	x		x					x			
PARTICIPANT 12			x	x	x						
PARTICIPANT 13			x			x					
PARTICIPANT 14			x								
PARTICIPANT 15			x								
PARTICIPANT 16			x		x						
PARTICIPANT 17			x		x						
PARTICIPANT 18	x		x			x					
PARTICIPANT 19			x				x				
PARTICIPANT 20			x								
PARTICIPANT 21		x	x								
PARTICIPANT 22			x				x				
PARTICIPANT 23			x		x						
PARTICIPANT 24			x								
PARTICIPANT 25	x		x			x					
PARTICIPANT 26			x								
PARTICIPANT 27	x			x	x	x					
PARTICIPANT 28		x	x		x						
PARTICIPANT 29		x	x								
PARTICIPANT 30			x				x				
PARTICIPANT 31			x			x	x				
PARTICIPANT 32			x	x	x						

Table G2. Reading Operations Agreed by Raters for the SUMONLY Task

READING OP. FOR SUMMARY T.	RO1	RO2	RO3	RO4	RO5	RO6	RO7	RO8	RO9	RO10	RO11
PARTICIPANT 1	x					x				x	
PARTICIPANT 2	x					x		x		x	x
PARTICIPANT 3	x					x	x				
PARTICIPANT 4	x					x			x		x
PARTICIPANT 5	x					x				x	
PARTICIPANT 6	x					x					
PARTICIPANT 7	x					x		x			
PARTICIPANT 8	x					x	x			x	x
PARTICIPANT 9	x					x					
PARTICIPANT 10	x					x				x	
PARTICIPANT 11	x					x	x	x	x		x
PARTICIPANT 12	x					x	x				
PARTICIPANT 13	x					x					
PARTICIPANT 14	x					x		x	x	x	
PARTICIPANT 15	x					x					
PARTICIPANT 16	x					x					
PARTICIPANT 17	x					x				x	
PARTICIPANT 18	x					x	x	x			x
PARTICIPANT 19	x					x	x				
PARTICIPANT 20	x					x	x				
PARTICIPANT 21		x						x	x	x	x
PARTICIPANT 22	x					x					
PARTICIPANT 23	x					x					
PARTICIPANT 24	x					x	x				
PARTICIPANT 25	x					x	x	x		x	x
PARTICIPANT 26	x					x					
PARTICIPANT 27	x					x		x		x	
PARTICIPANT 28	x					x					
PARTICIPANT 29	x					x				x	
PARTICIPANT 30	x					x	x	x	x		
PARTICIPANT 31	x					x				x	x
PARTICIPANT 32	x					x	x	x			

APPENDIX H

PARTICIPANTS' SCORES

Table H. Participants' Scores from all Tasks

P.NO	TOTAL SCORE LIST		
	MC	MCSUM	SUMONLY
1	50	57.1	42.8
2	62.5	85.7	42.8
3	87.5	57.1	28.5
4	75	42.8	57.1
5	87.5	28.5	28.5
6	100	100	71.4
7	75	71.4	42.8
8	37.5	100	57.1
9	87.5	42.8	28.5
10	87.5	71.4	28.5
11	25	57.1	28.5
12	62.5	42.8	71.4
13	75	100	100
14	87.5	100	85.7
15	75	71.4	42.8
16	75	42.8	57.1
17	62.5	85.7	42.8
18	62.5	42.8	71.4
19	50	42.8	14.2
20	87.5	85.7	71.4
21	25	57.1	71.4
22	62.5	85.7	57.1
23	37.5	71.4	42.8
24	62.5	57.1	42.8
25	62.5	57.1	14.2
26	40	42.8	28.5
27	37.5	28.5	14.2
28	50	71.4	85.7
29	50	71.4	28.5
30	62.5	71.4	71.4
31	50	71.4	28.5
32	75	85.7	57.1

REFERENCES

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK; New York, NY, USA; Cambridge University Press.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye tracking. *Language Testing*, *30*(3), 1–25.

Benson, J., & Crocker, L. (1979). The effects of item format and reading ability on objective test performance: A question of validity. *Educational and Psychological Measurement*, *39(2),* 381-387.

Bensoussan, M., & Kreindler, I. (1990). Improving advanced reading comprehension in a foreign language: Summaries vs. short-answer questions. *Journal of Research in Reading, 13(1)*, 55-68.

Bernhardt, E. B. (1983). Three approaches to reading comprehension in intermediate German. *Modern Language Journal, 67*, 111-115.

Brown, A.L., & Day, J.D. (1983). Macrorules for summarizing texts. The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*, 1-14.

Carver, R.P. (1997) Reading for one second, one minute, or one year from the perspective of rauding theory. *Scientific studies of reading*, *1*, 3-43.

Cerdan, R., Vidal-Abarca, E., Martinez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Language and instruction*, *19*, 13-27.

Chaka, N., & Booi-Ncetani (2015). An investigation into the English reading comprehension of Grade 10 English first additional language learners at a senior secondary school. *Reading & Writing*; *6*, 7.

Cohen, A. (1984). On Taking Language Tests: What the Students Report. *Language Testing*. *1*. 10.

Cohen, A.D. (1994). English for academic purposes in Brazil: the use of summary tasks. In C. Hill & K. Parry (Eds), *From testing to assessment: English as an international language*. London: Longman, 174-204.

Cohen, A. D., & Upton, T. (2007). `I want to go back to the text': Response strategies on the reading subtest of the new TOEFL(R). *Language Testing, 24(2)*, 209–250.

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

Deane, P., Sheehan, K., Sabatini, J., & Futagi, S. (2006). Differences in text structure and its implications for assessment of struggling readers. *Scientific Studies of Reading, 10(3*), 257-276.

Enright, M.K, Grabe, W., Koda, K., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 Reading Framework: A working paper*. TOEFL monograph series, MS-17, Princeton, NJ: ETS.

Fisher, A.S.H. (2016). Students' reading techniques difficulties in recount text, *Journal of English and education, 4(2),* 1-12.

Geva, E., Wade-Woolley, L., & Shany, M. (1997). Development of reading efficiency in first and second language. *Scientific Studies of Reading, 1(2)*, 119-144.

Gil, L., Braten, I., Vidal-Abarca, E., & Stromso, H.I. (2009). Summary versus argument tasks when working with multiple documents: Which is better for whom?. *Contemporary Educational Psychology, 35*, 157-173.

Goodman, K.S. (1967). Reading: a psycholinguistic guessing game?. *Journal of the Reading Specialist, 6*, 126-135.

Gough, P.B. (1972). One second of reading. *Visible language, 6*, 290-320.

Gove, M.K. (1983). Clarifying teachers' beliefs about reading. *The Reading Teacher, 37*, 261-268.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25*, 375-406.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* New York: Cambridge University Press.

Guthrie, J.T., & Kirsch, I.S. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology, 79*, 220-227.

Haladyna, T.M., & Downing, S.M. (2009). A Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education, 2(1)*, 37-50.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing, 26(2)*, 219–244.

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

Khalifa, H., & Weir, C. (2009). *Examining reading: research and practice in assessing second language reading.* Studies in Language Testing, *29*. Cambridge, England: Cambridge University Press.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85(5),* 363-394.

Kintsch, W. (2004) The Construction-Integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.) *Theoretical Models and Processes of Reading.* 5th Edition, International Reading Association.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49(4)*, 294-303.

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

Kintsch, W., & Kintsch, E. (2005). Comprehension. In Paris, S. G. and Stahl, S. A. (eds.) *Children's Reading Comprehension and Assessment*, 71–92. Mahwah, New Jersey: Erlbaum.

Kintsch, W., & van Dijk, T.A. (1983). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Kirsch, I.S., & Mosenthal, P. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, *25,* 5-30.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19(2)*, 193-220.

Kurz, T.B. (1999). *A review of scoring algorithms for multiple choice tests*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX. Retrieved from https://files.eric.ed.gov/fulltext/ED428076.pdf

Lau, P.N.K., Lau, S.H, Hong, K.S., & Usop, H. (2011) Guessing, partial knowledge, and misconceptions in multiple choice tests. *Educational Technology & Society*, *14*, 99-110.

Lee, J.F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in second language acquisition*, *8*, 201-212.

Lee, S. (2009). Topic congruence and topic interest: How do they affect second language reading comprehension?. *Reading in a Foreign Language*, *21*, 159-178.

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.

Leong, C. K., Tse, S. K., Loh, K. Y., & Hau, K. T. (2008). Text comprehension in Chinese children: Relative contribution of verbal working memory, pseudoword reading, rapid automatized naming, and onset-rime phonological segmentation. *Journal of Educational Psychology, 100*(1), 135-149.

Lim, H. J. (2014). Exploring the validity evidence of the TOEFL IBT reading test from a cognitive perspective. Unpublished PhD Thesis. Michigan State University.

Linderholm, T. (2006). Reading with purpose. *Journal of College Reading and Learning*, *36(2)*, 70-80.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*, 778-784.

Linderholm, T., & Wilde, A. (2010). College students' beliefs about comprehension when reading for different purposes. *Journal of College Reading and Learning*. *40*, 7.

Lo, L., Ho, C., Wong, Y., Chan, D., & Chung, K. (2016). Understanding the microstructure and macrostructure of passages among Chinese elementary school children. *Journal of Psycholinguistic Research*, *45*, 1287-1300.

Marzec-Stawiarska, M. (2015). The influence of summary writing on the development of reading skills in a foreign language. *System*, *59*, 90-99.

Magliano JP, Millis KK, Ozuru Y., & McNamara DS. (2007). A Multidimensional Framework to Evaluate Reading Assessment Tools. In: McNamara DS, editor. *Reading Comprehension Strategies: Theories, interventions, and technologies*. Mahwah, NJ: Erlbaum.

McCormick, T. W. (1988). *Theories of reading in dialogue: An interdisciplinary study*. New York: University Press of America.

McNamara, D. S. (2007). *Reading Comprehension Strategies: Theories, interventions, and technologies*. Mahwah, NJ: Erlbaum.

Mok, W.S.Y., & Chan, W.W.L. (2016). How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? *Instructional Science*, *44(6)*, 567-581.

Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. Language Teaching Methodology Series. New York: Prentice Hall.

Parry, K. (1991). Building a Vocabulary Through Academic Reading. *TESOL Quarterly*, *25*, 629–653.

Pang, J. (2008). Research on good and poor reader characteristics. *Reading in a Foreign Language*, *20*, 1-18.

Perfetti, C. A., & Roth, S. (1980). Some of the interactive processes in reading and their role in reading skill. In Lesgold, A, Perfetti, C. (Eds), *Interactive processes in reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pressley, G.M. (2002). Metacognition and self-regulated comprehension. In Farstrup, A.E., & Samuels, S.J. (Eds.), *What research has to say about reading instruction*. Newark, DE: International Reading Association.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity. *Language Learning*, *57*, 155-199.

Rauch, D. P., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, *52(4)*, 354.

Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, *23(4)*, 441–474.

McVee, M.B., Dunsmore, K., & Gavelek, J.R. (2005). Schema Theory Revisited. *Review of Educational Research*, *75*, 531-566.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *1(2)*, 147-170.

Squire, J.R. (1983). Composing and Comprehending: Two Sides of the Same Basic Process. *Language Arts*, *60*, 581–89.

Taylor, L. (2013) Testing reading through summary: investigating summary completion tasks for assessing reading comprehension ability. *Studies in Language Testing, 39*. Cambridge, England, UCLES/Cambridge University Press.

Urquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London, UK: Longman.

Ünaldı, A. (2004). Componentiality of the reading construct: Construct validation of the reading subskills of the Boğaziçi University English Proficiency Test. Unpublished PhD Thesis. Faculty of Education: Boğaziçi University.

van Dijk, T.A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, *22*, 293-301.

Weir, C, Hawkey, R, Green, T., & Devi, S. (2009). The cognitive processes underlying the academic reading construct as measured by IELTS. *IELTS Research Reports*, *9*, 157–189, British Council, London and IELTS Australia, Canberra.

Weir, C., Hawkey, R., Green, A., Unaldi, A., & Devi, S. (2012) The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university. In Taylor, L., & Weir, C. (eds.), *IELTS Collected Papers 2: Research in Reading and Listening Assessment, 2,* 37–120, Cambridge, England, Cambridge University Press.

Wolf, D.F. (1991). The effects of task, language of assessment, and target language experience on foreign language learners' performance on reading comprehension tests. Dissertation, University of Illinois, ProQuest Dissertations and Theses.