Title Investigating reading for academic purposes: Sentence, text and multiple texts

Name Aylin Unaldi

3403602906

INVESTIGATING READING FOR ACADEMIC PURPOSES:

SENTENCE, TEXT AND MULTIPLE TEXTS

by

AYLIN UNALDI

A thesis submitted for the degree of Doctor of Philosophy
of the University of Bedfordshire

December 2010

INVESTIGATING READING FOR ACADEMIC PURPOSES:

SENTENCE, TEXT AND MULTIPLE TEXTS

AYLIN UNALDI

ABSTRACT

This study examines the nature of reading in academic environments and suggest ways for a more appropriate assessment of it. Research studies show that reading in academic settings is a complex knowledge management process in which information is selected, combined and organised from not a single, isolated text but from multiple information sources.

This study initially gathered evidence from students studying at a British university on their perceived and observed reading purposes and processes in three studies; a large scale questionnaire, longitudinal reading diary study and finally individual interviews in order both to establish whether the prominent reading skills used by them were as put forth in the studies on academic reading, and to examine in detail the actual cognitive processes (reading operations) used in reading for academic purposes.

The study draws on the reading theories that explain reading comprehension and focuses specifically on different levels of careful reading such as sentence, text and multiple texts in order to explicate that increasingly more complex cognitive processes explain higher levels of reading comprehension.

Building on the findings from the three initial studies, it is suggested that reading tests of English for Academic Purposes (EAP) should involve not only local level comprehension questions but also reading tasks at text and multiple texts levels. For this aim, taking the Khalifa and Weir (2009) framework as the basis, cognitive processes extracted from the theories defining each level of reading, and contextual features extracted through the analysis of university course books were combined to form the test specifications for each level of careful reading and sample tests assessing careful reading at sentence, text and intertextual levels were designed. Statistical findings confirmed the differential nature of the three levels of careful reading; however, the expected difficulty continuum could not be observed among the tests. Possible reasons underlying this are discussed, suggestions on reading tasks that might operationalise text level reading more efficiently and intertextual level reading more extensively are made and additional components of intertextual reading are offered for the Khalifa and Weir (2009) reading framework. The implications of the findings for the teaching and assessment of English for Academic Purposes are also discussed.

To my son Deniz, who has made life meaningful to me.

DECLARATION

The studies described in this thesis relate to the research focus of CRELLA, the host institution. The questionnaire study was designed and executed by CRELLA research team; Cyril Weir, Roger Hawkey, Anthony Green, Sarojani Devi, with the researcher herself entirely responsible for developing and validating the questions identifying careful reading activities. This study was published in *IELTS Research Reports* Volume 9 (Weir, et al., 2009). Anthony Green collaborated with the researcher in the development of the Contextual Validity Proforma and the text analysis. This study was published in *Language Testing* (Green et al., 2010).

Otherwise, I declare that this thesis is my own unaided work. It is being submitted for the degree of PhD at the University of Bedfordshire.

It has not been submitted before for any degree or examination in any other University.

Name of candidate: Aylin Unaldı Signature:.........................

Date: December 2010

ACKNOWLEDGMENTS

If it were not unprecedented to write a novella at the acknowledgement section of a PhD thesis, I had one ready. At the age of 40 and with a one-year old child, if a woman sets out for a second PhD in a foreign country, she will definitely have a story to tell. Well, women are strong, as we all know. We can survive many difficulties. And a PhD starts and eventually finishes but what remains behind creeps into our lives and changes us as a person gradually and silently.

My colleagues at CRELLA took me from one place in my life and carried me to a much higher place. Not only did they train me as a researcher, they taught me the meaning of patience, discipline, collaboration and trust. More importantly, they showed me that all these could be done in an atmosphere of friendship. They told me that my best can be good enough.

Professor Cyril Weir, I will not thank you. Thanking is not enough as I owe you everything. I cried on your shoulder. I borrowed money from you. I used you as a free therapist. When I was at the edge, you pulled me back. You were a mentor, a brother and a friend to me. It was such a pleasure to talk about academic matters with you around dinner tables and it was such a big pride to be your student. I will miss it a lot.

Dr. Tony Green, the man of golden silence. If you were not sitting right across me there in the office, this PhD would be impossible. You patiently answered all my questions that came across every six seconds. You were always clear, helpful and supportive and you never complained. It felt great to be working with such an intelligent man as you.

Professor Roger Hawkey, when you were at CRELLA, the days were brighter. The encouragement I received from you, your ever-positive attitude and accurate recommendations were invaluable to me.

Professor Liz Hamp-Lyons, in the brief instances I could work with you, I wished I had more opportunity to benefit from you. Your existence made a difference in CRELLA.

My friend, Dr. Fumiyo Nakatsuhara, I heartily thank you for your sincere friendship and unconditional help you gave me. I am sure you have a long and successful Language Testing path in front of you that you will walk with CRELLA.

My dear friend, Betigul Guney, you accompanied me at difficult times, advised, supported and even fed me. I would not be able to type these words so comfortably now if you were not here to help me.

I am also grateful to all Putteridge Bury staff, who made the place a home to me. My friend Nick Gardner has a special place among them. My mother, as always, was behind the scene to support me. And above everything, there was one little person with me all through this adventure. When he held my hand, I took in all the energy in the universe and felt capable of doing anything. If there was one reason why I wanted to be successful, it was because I was his mother. Dear little buddy, thank you for being in my life.

Finally, I should express my gratitude to the University of Bedfordshire, Business and Management Research Institute as they fully sponsored this study. Thanks should also go the TOEFL Grants and Awards Committee, Educational Testing Service. They gave a very timely and helpful support to this study through The Small Grants for Doctoral Research in Second or Foreign Language Assessment scheme.

LIST OF CONTENTS

LIST OF TABLES

Table:

LIST OF FIGURES

Figure:

ABBREVIATIONS

ACW	Average Character per Word
AIID	Alpha if Item Deleted
ANOVA	Analysis of Variance
ASP	Average Sentence per Paragraph
AWS	Average Word per Sentence
AWL	Academic Word List
BNC	British National Corpus
CAE	Certificate in Advanced English
CEFR	Common European Framework of Reference
CITC	Corrected Item Total Correlations
CPE	Certificate of Proficiency in English
EAL	English as an Additional Language
EAP	English for Academic Purposes
EFL	English as a Foreign Language
EL1	English as a First Language
ESOL	English for Speakers of Other Languages
EU	European Union
FCE	First Certificate in English
FKGL	Flesh-Kincaid Grade Level
FRE	Flesh Reading Ease Score
HESA	The Higher Education Statistic Agency
IELTS	International English Language Testing System

ID	Item Discrimination
IDP	Item Discrimination Pattern
IF	Item Facility
ISE	Integrated Skills in English
KMO	Kaiser-Meyer-Olkin
L1	First Language
L2	Second Language
LD	Lexical Density
LSA	Latent Semantic Analysis
PAF	Principle Axis Factoring
PCA	Principle Component Analysis
STTR	Standardised Type Token Ratio
TOEFLibt	Test of English as a Foreign Language – Internet Based
TTR	Type Token Ratio
UCLES	University of Cambridge Local Examinations Syndicate

CHAPTER 1

INTRODUCTION

1.1 Background to the Study

In academic settings, a substantial portion of knowledge is acquired through written documents, and the ability to process a text or texts efficiently with sufficient comprehension is a sine-qua-non (Carson, 2001). Higher education institutions require a high degree of literacy on the part of students on entry (Snow, 2002) and this includes assumptions on students' ability to comprehend complex texts and being able to communicate ideas through written language.

Although reading comprehension is an area of research that has perhaps attracted more attention than any other language skill and has been researched from a multitude of aspects, the interest in the careful reading skill - slow, incremental, non-selective reading - at higher levels such as at the text and intertextual levels is relatively recent and narrow in scope, and less than scarce in second language research. Most attention in attempts to understand the nature of the reading processes has been devoted to the analyses of lower level processes in careful reading such as word recognition, lexical access, syntactic parsing and sentence comprehension (i.e. Just and Carpenter 1980, 1987, 1992; Carver 1997, 1998; Perfetti 1991i; Rayner and Pollatsek, 1989; Rumelhart and McClelland et al. 1986).

A few models such as Kintsch and van Dijk (1978) and van Dijk and Kintsch's (1983) 'text model comprehension' place greater importance on propositional integration across sentences and the overall discourse processing level (Barnett, 1989, 27). Researchers in the field of discourse studies explain reading comprehension in a similar fashion. For example, for Gernsbacher (1990), Goldman and Rakestraw (2000) and Meyer (1999), the goal of comprehension is building cohesive mental representations or structures of a text. Discourse models of reading comprehension emphasise the importance of accurate processing of the hierarchical relations of the elements of a text – relations of sentential elements and relations between sentences as well as relations between groups of sentences, i.e. paragraph structures and functions – in arriving at an organised and coherent mental representation of the whole text.

On the other hand, there is little research on, but an emerging interest in what goes on beyond text comprehension when readers read multiple texts. Among reading models, one that moves beyond the conceptualisation of comprehension within the limits of a text (text base and situational model) is the 'documents model' (Perfetti, 1997 and Perfetti et al., 1999). Perfetti et al. (1999) claim that readers can construct spatial and non-propositional representations that are not based on the explicit content of the text but as derivative of it. Britt and Sommer (2004) state that learning by reading multiple texts involves the same processes and structures as single-text reading but require additional skills for coordinating separate and sometimes discrepant presentations and for handling document level information. Since texts are not normally written to be read in conjunction with other texts, they lack explicit links to facilitate integration of information across texts, and the demands on the reader to form a macrostructure are higher than when reading a single text with intratextual coherence.

The growing need to explicate the unique characteristics of reading multiple texts has been stimulated by the attempts to reconceptualise 'academic literacy' and a renewed interest in 'integrated language skills'. This interest has recently been reflected in several high-stakes English for Academic Purposes (EAP) language tests, however, with inadequate support from theory and research.

On the other hand, it has been well established in the testing field that the link between the theory and the test should be built through the development of test specifications, which should specify the types of test tasks and of the expected responses in terms of the contextual features and cognitive processes that delimit them. Therefore, establishing the theory that explains particulars of the construct in question, defining the test specifications that would transform the contextual and cognitive premises of the theory to operational definitions, and building tasks and defining response features according to the test specifications are the first steps of test development and validation (Messick 1989, Bachman 1990, Alderson, 2000, Weir 2005).

Hence, there is still a need in the area of language testing to establish the theoretical basis for different types of careful reading, especially careful reading at the multiple text (intertextual) level, and to investigate the different cognitive and contextual features that define the types of careful reading to inform the design and use of language tests.

1.2 Aims of the Research

The present study aims at a closer look into reading processes in academic settings with a focus on careful reading processes whereby readers process texts incrementally and non-selectively. It aims to pull together suggestions from theoretical models that explain several aspects of reading comprehension individually, and to evaluate evidence for the existence of different types of careful reading by analysing the real-life reading processes of students and their performance on a test which is constructed to measure the different types of careful reading. It should therefore provide empirical support for the premise of a more comprehensive careful reading model.

To recapitulate the aims of the study briefly, it will focus on:

- the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution in the UK, and the cognitive and contextual features that underlie them,
- how students perform on tests constructed to measure different types of careful reading,
- the cognitive processes that underlie the different types of careful reading engaged in by students in these tests.

Three research questions have been formulated for the investigation of these issues:

Research Question 1) What are the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution?

Research Question 2) Do test takers score differently on tasks operationalising three careful reading types at sentence, text and multiple text levels?

Research Question 3) Is cognitive processing different in the various types of careful reading?

1.3 Overview of Methodology

In order to investigate the research questions, several studies have been designed and these will be detailed in the methodology chapter of the thesis. However, an outline of the particular studies might be helpful.

The investigation of Research Question 1 consists of four stages:

1. Literature review on types of L2 academic reading
2. Questionnaire study to determine the academic reading activities and cognitive and contextual parameters shaping first year academic reading at the university level
3. A longitudinal reading diary study for a closer look into the reading activities of university students while they are reading for writing purposes.
4. Follow-up interviews with the diary study participants to investigate the details on the cognitive processes and strategies that the readers use when they are reading multiple texts for writing purposes.

The investigation of the second research question entails the development of reading tests measuring different levels of careful reading. The test development involves three preliminary studies that pave the way to the tests. These are document analysis, context validity proforma and development of test specifications. Firstly, major EAP tests will be analysed for the investigation of operationalisation of careful reading construct. The operationalisation of careful reading tasks in EAP tests will demonstrate whether careful reading is conceived as having different types or not. Context validity proforma study will reveal the textual features that delimit academic reading at the university level and will guide the development of both the test specifications and the selection of test texts. The next step is the development of test specifications and test tasks. This part of the study will reveal a detailed procedure of developing test specifications and test tasks in which conceptualisation of multi-level careful reading has been instantiated. As these studies involve important details of the test development process, they will be reported in the methodology chapter in full. The investigation of the second research question will involve a qualitative pilot study on the tests and the statistical analysis of the test data. Verbal protocols of test takers will be used to analyse the congruence between the test tasks and test specifications *a priori*, and the processes test takers use in responding the test items will be investigated. This will help improvement of the tests where necessary. Following this, the tests will be administered to a large group of test takers and the data will be analysed statistically.

Research Question 3 will be investigated by the analysis of the cognitive validity proforma data the test takers will respond to as they have taken the test. The results will be discussed in combination with the findings from verbal protocol study mentioned above.

1.4 Overview of the Study

Following this introductory chapter, the thesis will present in Chapter 2 a review of the literature on the issues concerning academic study and the types of L2 reading, the levels of cognitive processing in careful reading, and reading models and testing of English for academic purposes.

In Chapter 3, the methods of investigation in relation to each research question will be described. This chapter will inform the reader on the details of the development of the questionnaire, the reading diary, the context validity proforma, the cognitive validity proforma, the test specifications and the test tasks and related studies. The chapter will also introduce the details of the statistical procedures used in the analyses of the research questions.

Chapter 4 will present the results and discussion concerning the nature of careful reading process. Finally, in Chapter 5 a summary of the findings will be presented, the implications of the findings will be discussed, and the limitations of the study will be considered with suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Efficiency and accuracy in reading are deemed to be very important to success in academic settings (Carson, 2001). Readers in academic environments are expected to acquire new knowledge and learn new concepts by processing complex texts and applying textual knowledge appropriately (Snow, 2002). In tests of target language reading ability, as in any other type of test, an appropriate theory of the ability in question should be operationalised in order to generalise the inferences based on test scores beyond the test situation (Alderson, 2000). Khalifa and Weir (2009) underline the importance of theory in terms of cognitive processing and contextual parameters as the basis of test development. Therefore, the ability – the construct – in question should be defined 'in terms of a theory that specifies how it relates to other constructs and observable performance' (Bachman 1990, 225).

In several frameworks of test development and validation (e.g. Messick 1989, Bachman 1990, Weir 2005), it is emphasised that the link between the theory and the test should be built through the development of test specifications which should specify the types of test tasks and the expected responses in terms of the contextual features and cognitive processes that delimit them. As mentioned above, establishing the theory that explains particulars of the construct in question, defining the test specifications that would transform the contextual and cognitive premises of the theory to operational definitions, and building tasks and defining response features according to the test specifications are the first steps of test development and validation.

The nature of reading has frequently been defined through taxonomies which divide the reading construct into skills and sub-skills as they relate to reading purposes and processes (Carver, 1997; Grabe and Stoller, 2002; Munby, 1978; Urquhart and Weir; 1998). There is usually a divide in the taxonomies between the types of reading that require quick and strategic reading for the purpose of searching for information as opposed to slower, more detailed reading that involves incremental processing of a text for the purpose of learning from the text. Urquhart and Weir (1998) refer to the former as 'expeditious reading', and the latter, 'careful reading'[1]. This study will focus on careful reading in a larger context of academic reading and will identify careful reading skills that need to be assessed in EAP tests, suggest ways for their operationalisation in the tests, and should eventually help broaden our understanding of assessment of careful reading in EAP settings. The initial step will be the

[1] See Section 2.2

investigation of careful reading processes in academic environments by reviewing theoretical models that explain several aspects of reading. Secondly, the real-life reading processes of students and their performance on a test which is constructed to measure the different types of careful reading will be evaluated to provide evidence for the existence of different types of careful reading.

In this chapter, a theoretical basis for describing the nature of academic reading in English-medium university settings will be proposed through a discussion of sample studies with a view to determining the importance of careful reading types in academic environments. Careful reading at sentence, text and intertextual levels will then be focused from a theoretical perspective to inform the analysis of the different cognitive processes involved in each of them[2]. Finally, two reading frameworks which cover explicitly all three types of careful reading and which were developed with academic reading assessment in mind – Enright et al. (2000) and Khalifa and Weir (2009) – will be reviewed critically for the suitability of the theoretical constructs in terms of informing the development of an EAP reading test assessing different levels of careful reading.

[2] In this study, the term cognitive process is used as such processes are defined in the language teaching and assessment field; the processes are reading operations that a reader performs when dealing with a text.

2.2 Academic Study and Types of Careful Reading

Our knowledge concerning the nature of reading in academic settings comes from research that investigates the texts and tasks assigned in university courses and to a much lesser extent, from the analysis of the reading processes of students. Strong suggestions on the place and the nature of reading also come from similar studies on academic writing.

One of the key studies that has investigated reading activities in university programs and informed the design of EAP reading instruction and language proficiency tests is Weir (1983), in which several careful reading skills as well as expeditious reading strategies were identified as important both for undergraduate and post graduate study. Urquhart and Weir (1998) reflect a refined form of Weir (1983) in a framework of types of reading that are shaped by the purposes and processes of reading encapsulating both local versus global and expeditious versus careful types of reading skills that define successful academic reading. In the framework, careful reading at the local level comprises ‘understanding syntactic structure of sentence and clause including lexical and/or grammatical cohesion and understanding lexis/deducing meaning of lexical items from morphology and context. Careful reading at the global level on the other hand involves reading carefully to establish accurate comprehension of the explicitly stated main ideas the author wishes to convey and propositional inferencing (see Figure 1). The successful acquisition and use of these reading skills have been identified as crucial for success in academic life and proposed as the premises for EAP teaching and testing (Urquhart and Weir, 1998).

Figure 2.1: Types of reading (Urquhart and Weir, 1998)

	Global level	Local level
Careful Reading	▪ Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences ▪ Making propositional inferences ▪ Establishing how ideas and details relate to each other in a whole text ▪ Establishing how ideas and details relate to each other across texts	▪ Establishing accurate comprehension of explicitly stated main idea or supporting details within a sentence ▪ Identifying lexis ▪ Understanding syntax
Expeditious Reading	▪ Skimming quickly to establish: discourse topic and main ideas, or structure of text, or relevance to needs ▪ Search reading to locate quickly and understand information relevant to predetermined needs	▪ Scanning to locate specific points of information

An alternative framework has been suggested by Enright et al (2000). Enright et al.'s (2000, 5-6) academic reading framework for TOEFLibt, designed on the conceptualisation of reading from 'reader purpose' perspective, formulates four types of reading purpose;

.reading to find information (search reading)

.reading for basic comprehension

.reading to learn (building a mental model)

.reading to integrate information across multiple texts (documents model).

The framework will be discussed in more detail in Section 2.4.

Based on the theoretical models proposed in the TOEFL 2000 framework monographs (http://www.ets.org/toefl/research/monograph_series), Rosenfeld et al (2001) identified language proficiency tasks that need to be carried out competently for success in academic life. Among all the tasks rated as very important both by faculty and students, three careful reading purposes were present:

Basic comprehension: determining the basic theme (main idea) of a passage; reading and understanding written instructions/directions concerning classroom assignments and/or examinations.

*Reading to learn: r*eading text material with sufficient care and comprehension to remember major ideas and answering written questions later when the text is no longer present; reading text material with sufficient care and comprehension to remember major ideas.

Reading to integrate: comparing and contrasting ideas in a single text and/or across texts; synthesizing ideas in a single text and/or across texts.

A more recent framework of academic reading designed to unite the factors that bear direct relevance to the assessment of EAP reading ability and to guide test development and evaluation processes is Khalifa and Weir (2009). Khalifa and Weir's framework, which is going to be presented in more detail in Section 2.4, is an extension of Urquhart and Weir (1998) and it accommodates the cognitive processing required for different levels of reading starting from the word level and extending to the multiple text level. Briefly, it accounts for the reader's purpose in the selection of the type of reading to be engaged, and the knowledge base the reader brings to the reading process for the comprehension process to run smoothly. Contextual parameters, i.e., text and task features, performance conditions that affect reading comprehension and thus test performance are explicitly accounted for. Khalifa and Weir (ibid.) present a detailed analysis of the UCLES Main Suite Exams rather than an analysis of academic reading in university settings, however, it is the most up to date academic reading model with an explicit application of a reading model to reading tests. Besides, this framework was used as the basis of a study of the reading experiences of first year university students studying at an English university, analysing the academic reading needs of both home and overseas students from the perspective of the students themselves (Weir et al, 2009).

Another recent study on the academic reading requirements at the university level is Moore et al. (2010). Moore et al. (ibid.) comparing the reading requirements in IELTS test items and in university study show that reading practices in university settings are shaped not only by basic comprehension limited to relatively small textual units but also by more global reading activities that involve multiple sources and view points. Tasks that require engagement in multiple texts and integration of reading activities into some related writing activity such as summary tasks and essay tasks are prevalent in academic settings. Moore et al. (ibid.) also report that some academic informants, lecturers, in the study noted that locating, selecting and evaluating information sources is a considerable challenge for students, especially because of the vast amount of online information available to them. The lecturers underlined the necessity of an 'information literacy' approach in their courses, in which students are made aware of such matters as the context of text production, authorship, communicative purpose and the reliability of sources.

As well as above studies that focus on academic reading, studies aiming at analysing academic writing demands in university settings have important suggestions for understanding the nature of academic reading.

Horowitz (1986) for example, identifies seven types of academic writing tasks: summary of/reaction to a reading, annotated bibliography, report on a specified participatory experience, connection of theory and data, case study, synthesis of multiple sources, and research project, all of which are heavily reading embedded. He suggests that the pedagogy on academic writing should include 'selecting data which is relevant to a question or issue from a source or sources, reorganizing the data in response to the given question or issue, and encoding the data into academic English' (p. 456). The first step of academic writing is therefore selecting relevant information especially among multiple sources, a process which involves distinguishing what is useful from what is not among complementing, overlapping or contradicting information.

Hale at al. (1996), though their categorisation is inconclusive in many respects, identified tasks such as the library research paper (that require information search and incorporation of information in an essay), summarisation tasks (that might require the condensation of information from a single or multiple sources), and case studies and writing a plan or proposal, (both of which necessitated description of an approach – possibly theoretical or methodological. Accordingly, Cumming et al. (2000) stress the importance of accounting for the more integrated nature of writing tasks in authentic academic contexts in TOEFLibt.

Therefore, it can be claimed that in academic settings, reading (and thus learning) rarely depends on processing of a single text let alone processing of isolated sentences as in many EAP reading tests. Especially in higher education institutions, students are expected to form a disciplinary understanding by accumulating knowledge through an on-going reading of related texts. In assignment preparation, for instance, it is important for a student to use several sources and approach the information presented in the texts critically (Weir, 1983; Jordan, 1998; Carson, 2001; Kroll, 1990).

Rather than addressing the separate contributions of reading and writing skills to academic success, certain researchers have proposed frameworks of integrated reading and writing to explain the development of advanced academic abilities. In a constructivist reading-writing framework suggested by Spivey and King (1989, 1994)[3], the reading to

[3] See also Spivey (1990 and 1991).

write process is conceptualised as 'discourse synthesis' in which readers select, organise and connect information from multiple source texts as they compose their own texts. Reader characteristics such as prior knowledge and language ability, and task demands such as purpose and audience shape the reading processes and affect the constructions to be made while the readers synthesise information from the sources and build links between related ideas, organising them in a new text. Spivey and King (1994) state that this is a developmental process as proficient readers are more sensitive to intertextually important information and are more able to produce integrated structures with elaborated content.

Carson (2001), in her analysis of reading and writing tasks in academic contexts, stresses that in academic settings integration across skills is essential. Reading is especially important in the preparation stage of academic tasks in which students need to identify relevant information, analyse meaning relationships between the pieces of information and organise information and interpret the meaning of texts. Carson (ibid.) underlines the point that studies on integrated reading and writing tend to focus on the writing aspect, and Grabe (2001; 2004) points out the need for a reading-writing theory that can account for the coordinated processing demands of reading and writing together. Grabe (2001, 18) underlines that 'many discussions on reading-writing interactions particularly in L2 contexts, focus primarily on writing issues, seeing reading as a springboard to writing tasks and learning for writing.' He (ibid., 18) suggests that reading issues, as well as those of writing, should be given due consideration in any serious analysis as 'the overlap between reading and writing processing and abilities is not simply a matter of conventional recognition of two sides of literacy.'

As evident from the discussion above, in academic settings, reading is a multi-layered, multi-faceted process that takes place purposefully in conjunction with other academic tasks. Basic comprehension at local levels can neither sufficiently explain nor secure academic success. The present study aims at focusing on reading at different textual levels with special emphasis on the multiple text (intertextual) level so that reading at intertextual level can be better understood as a reading skill per se. On this basis, intertextual reading can be approached as a skill that can be operationalised and assessed in EAP reading tests. In order to understand the cognitive processes unique to sentence, text and multiple text levels, the next section will focus on reading theories that explain reading at different levels.

2.3 Levels of Cognitive Processing in Careful Reading

Conceptualisation of reading comprehension depends largely on the interests of the researchers working in a specific field of study. By and large, cognitive psychologists primarily focus on reading words, phrases or a sentence. The next level, text level reading, is within the scope of several fields, such as education, cognitive psychology and discourse linguistics. There is a very recent interest in the reading processes beyond the text level, i.e. reading at multiple text level as a distinct and higher level of reading comprehension. The existence of a variety of approaches to the explanation of careful reading is itself suggestive of the possibility that careful reading at different levels may involve different cognitive processes. Therefore, it is important to look at some of the widely accepted reading models explicating the comprehension process and putting primacy on different levels of text.

2.3.1 Sentence Comprehension

There is a multitude of research in the area of reading comprehension both in the first and second/foreign language. However, most attention in attempts to understand the nature of the reading processes has been devoted to the analyses of lower level processes in careful reading such as word recognition, lexical access, syntactic parsing and sentence comprehension. In reading models that focus on lower level processing, reading is a process in which lower units are analysed and gradually added to higher units until the meaning is constructed through the application of syntactic and semantic rules. Comprehension takes place after this series of operations are complete with little influence from general world knowledge, contextual information or higher order processing strategies (Barnett, 1989; Rayner and Pollatsek, 1989, Grabe and Stoller, 2002). For example in Just and Carpenter's (1980, 1987, 1992) model, the prominent level in reading is considered to be the lexical level, including encoding the printed word and accessing its meaning in a mental dictionary (lexical access). Just and Carpenter (1980) show that time spent on a lexical item is directly related to the amount of time needed to process that word. Readers make longer pauses at points where processing loads are greater (e.g. content words, important clauses, ends of sentences). The following five processes are suggested as the basic processes of reading:

1. seeing the next word and extracting its physical features
2. seeing the word as a word and comparing it to the mental lexicon
3. assigning a case (e.ig. nominative, objective) to the word
4. relating the word to the rest of the words
5. wrapping up the sentence when complete.

According to Just and Carpenter (1987), readers try to interpret each word of a text (immediacy of interpretation) as they read. Phrases and clauses are analysed at syntactic and semantic levels. In order to make sense of a text, the reader must construct a representation of the concepts and the situation to which the text is referring (referential representation). Component processes in reading are coordinated in time and can operate in parallel by using a common working memory. A production system, which is central to the model, operates on the contents of memory and triggers necessary production rules for the integration of a text structure or inserting new elements in working memory (in 'recognise-act' cycles). During a production cycle, contents of memory are simultaneously assessed through an interaction of productions and the production conditions (Stanovich, 1991 and 1996). Individual differences in language comprehension can be attributed to the variations in total amount of activation in memory, which is responsible for processing and storage (Just and Carpenter, 1992).

Rayner and Pollatsek (1989) also put a particular emphasis on lexical encoding. According to their model, the reading processing sequence begins during eye fixation with the initial encoding of the printed words after which lexical access takes place. Lexical access creates an auditory code (inner speech) and rules and analogies are activated automatically. This process may involve multiple lexical items simultaneously. As the lexical access is completed, the meaning of fixed words is integrated into an ongoing text representation in working memory.

Perfetti's (1991) Restricted Verbal Efficiency Model emphasises the importance of building a store of graphemically accessible words in learning to read. The focus of the Perfetti's (1985) research is the cognitive processes that take place during eye fixations that occur at fast rates (Carver, 1997).

Such models that place a primary emphasis on lexical access as the ones that have been discussed above do make reference to higher level processes (i.e. referential representation, Just and Carpenter, 1987). However, storage and retrieval of lexical items, syntactic and semantic information attached to them in the memory are of more importance. On the other hand, two of the influential models that are discussed below explicate language processing at a propositional as well as a lexical level.

Rumelhart et al.'s (1986 in Barnett, 1989) parallel distributed processing model focuses on the sentential level in explaining the mental processes involved in reading comprehension. They suggest that 'information processing takes place through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units'. These units represent hypotheses about words, syntactic elements, etc. Interconnections among units form the constraints known to exist between the hypotheses (Barnett, 1989; 27). Although this model accounts for the interaction and simultaneous operation among lower-level and higher level processes (i.e. semantic knowledge influencing word perception, word knowledge influencing syntax, etc.), reading comprehension is still explained by the processes of constructing representations of sentences.

Carver's rauding theory (1997 and 1998) emphasises cognitive processes such as lexical access, semantic encoding, sentence integrating, proposition integrating and idea remembering that will be used successively in differing rates and purposes of reading. However, comprehension of the thoughts within a sentence – lexical access, semantic encoding and sentence integrating – is the fundamental part of the rauding theory as this type of reading is seen as the most typical reading executed by adults reading texts at the level of difficulty of their ability level.

Early models in the field of second language reading emphasise lower level processes as well. For example, the psycholinguistic model proposed by Coady (1979) suggests that comprehension results from the interaction of conceptual abilities, background knowledge and process strategies. Individual process strategies are: phoneme-grapheme correspondences, graheme-morphophoneme correspondences, syllable-morpheme correspondences, syntactic information (deep and surface), lexical meaning and contextual meaning, cognitive strategies, and affective mobilisers. According to Coady's model, learners progress from reliance on concrete processing strategies (e.g. grapheme-phoneme correspondences) to more abstract strategies (e.g. contextual or lexical meaning) and shift processing strategies or change the balance between them to match different types of texts or to accomplish different goals.

2.3.2 Text Comprehension

The reading models mentioned above represent the cognitive psychology perspective in explaining reading comprehension. The studies done from this perspective primarily attempt to give a detailed account of the lower level cognitive processes such as word-recognition and syntactic parsing according to temporal sequence of reading processes through eye movement and computer on-line studies . They place primacy on lower-level processing for the efficiency of reading and envisage reading as a process of meaning making within the boundaries of a sentence. The sentence is considered to be the unit of comprehension, and comprehension of a text entails comprehension of sentences (Perfetti, 1997). However, in order to fully grasp the processes underlying a reading session in which a reader reads a full length text trying to make sense of and especially learn from it, these theorites are less than adequate in explaining higher level processes, i.e., how the information in the text is processed incrementally and how the meaning is extracted from the whole text (Rayner and Pollatsek; 1989).

A few models such as Kintsch and van Dijk (1978) and van Dijk and Kintsch's (1983) 'text model comprehension' place greater importance on propositional integration across sentences and the overall discourse processing level (Barnett, 1989, 27).

Kintsch and van Dijk (1978) in their model of text comprehension and production emphasise comprehension to the exclusion of word recognition, although they assume the latter must exist (Barnett, 1989, 27). The assumption is that the surface structure of a piece of discourse is interpreted as a set of propositions. Propositions (the meaning elements of a text, underlying semantic structures) become organised into a coherent whole (a text base) in differential retention according to their salience in the text. The semantic structure of texts can be described both at the local microstructure level (structure of the individual propositions and their relations) and at a more global macrostructure (discourse level), that is, by micropropositions and macropropositions. The formation of a coherent mental semantic text base (a discourse topic) involves a cyclical process maintained through macrorules based on referential coherence (argument overlap), and if referential coherence is scarce, on inference. However, the formation of the text base is constrained by limitations of working memory or buffer capacity. Macrooperators reduce information in a text base to its gist, that is, the theoretical macrostructure. These operations are under the control of schema (involving schematic structures of discourse; superstructures), which is a theoretical formulation of a comprehender's goal. Macrorules are the semantic mapping rules that organise propositions into appropriate levels (Kintsch and van Dijk, 1978, see also van Dijk, 1977). Van Dijk and Kintsch (1983, in Grabe, 1999 and Kintsch, 1988) particularly emphasise three levels of comprehension representation: 1) verbatim representation which decays rapidly, 2) conceptual text-based representation that is generated through the process described above, and 3) the situation model that incorporates the reader's schemata and affective states; a deeper level at which the text loses its individuality and its information content. It is at the last level that not only comprehension but learning takes place (Kintsch, 1994).

Kintsch (1988)[4] later revised the model to integrate lower level processes; a construction-integration model, in which the initial processing is strictly bottom-up. In this model, a text base is constructed from the linguistic input in a construction process. The text base is integrated with the comprehender's knowledge base (an associative network the nodes of which are concepts or propositions), while the text is integrated into a coherent whole through a spreading activation process, whose duty is to select the best interpretation through the control of inconsistencies and irrelevancies.

For Gernsbacher (1990) as well, the goal of comprehension is building cohesive mental representations or structures of a text. Readers develop mental structures by mapping information onto already existing structures when the incoming information coheres with previous information. When the incoming information is less coherent, the readers shift to initiate a new substructure. Information is stored in the form of representations comprising several branching substructures.

[4] See Kintsch (1998) for a more detailed account of the theory.

Other studies on text coherence and text structure (i. e. Meyer, 1999; Goldman and Rakestraw, 2000) emphasise the role of text as the ultimate unit in comprehension. Goldman and Rakestraw (ibid.) argue that building coherent mental representations of information in a text involves processing of individual words and phrases and how these relate to each other both with in the text and within the pre-existing knowledge-base. While the readers are building associative networks of nodes of information and links among those nodes, prior knowledge related to the text information contributes to the network. Readers therefore depend on text-driven processing (the use of content and organisation of the text) and knowledge-driven processing (prior knowledge on content as well as the reader's structural knowledge on syntax, paragraph organisation, discourse genre and rhetorical structures).

Several models of second language reading take into account text level processes as well. For example, Bernhardt's (1986) Constructivist Model of second language reading includes 'text-based' and 'extra text-based' components. According to Bernhardt, 'the reader recognises words and syntactic features, brings prior knowledge to the text, links the text elements together and thinks about how the reading process is working (metacognition)' (Barnett, 1989, 47). Bernhardt (1991) later revises her model to include three components: language, literacy and world knowledge. The language component includes word structure, word meaning, syntax and morphology. The literacy component involves the reader's preferred level of understanding, goal setting and comprehension monitoring. Higher levels of literacy will enable the reader to deploy different strategies. According to Bernhardt, literacy includes knowing how to approach a text, why one approaches it and what to do with it. The world knowledge, on the other hand, involves background knowledge a reader possesses and uses to facilitate comprehension.

Carrell (1988) also maintains that reading comprehension is characterised as involving an interaction of 'text-based' and 'knowledge based' processes (the latter indicating the reader's existing background) and the most efficient reading is a bidirectional combination of text-based and knowledge-based processes. Carrell and Eisterhold (1988, 79)[5] further revise the concept of background knowledge, drawing a distinction between 'formal schemata' (background knowledge of the formal, rhetorical, organisational structures of different type of texts) and 'content schemata' (background knowledge of the content area of a text), both of which are important information that a reader bring into the reading of a text.

Similarly, Grabe and Stoller (2002) emphasise the distinction between lower-level and higher-level processes activated when reading takes: Lower level processes such as lexical access, syntactic parsing, semantic proposition formation and working memory activation are considered to be

[5] See also Carrell (1990) and (1992).

automatic linguistic processes whereas higher-level processes such as text model comprehension, situation model of reader interpretation and executive control processes relate more to the use of background knowledge and inferencing skills. Even though Grabe and Stoller maintain that reading comprehension is 'balancing and coordinating many (of these) abilities in a very complex and rapid set of routines' (ibid., 29), they also stress that automatic processes are 'carried out in a bottom-up manner with little interference from other processing levels or knowledge sources'. For example, fluent word recognition or initial syntactic parsing does not require interaction from context or background information. When readers have problems at these levels, then structures are raised to the conscious level for the use of context and inferencing (ibid., 33). Grabe and Stoller, stressing the importance of different types of reading changing according to the reader purpose, note that the use of higher order skills might change according to the various purposes for reading; a reader might be using more top-down processing when skimming a text, for example.

In sum, discourse models of reading comprehension emphasise the importance of accurate processing of the hierarchical relations of the elements of a text – relations of sentential elements and relations between sentences as well as relations between groups of sentences, i.e. paragraph structures and functions – in arriving at an organised and coherent mental representation of the whole text. The information in the text is condensed to a base form in accordance with the reader's current goals and affective state. For the comprehension of a text, the readers resort to their background knowledge of several types - lexical, syntactic knowledge as well as prior information at higher levels on content and text structure. Moreover, discourse models of reading attempt to explain comprehension beyond the sentence by focusing on text characteristics such as genre-related features (development of discourse; organisation of information, establishment of coherence throughout a text and lexical choice in specific genres), which are irrelevant for the models mentioned in the previous section.

2.3.3 Intertextual Comprehension

In the most general understanding of reading comprehension, comprehension of a text is the ultimate level of analysis. The emerging interest, especially in the field of academic literacy, in what goes on beyond text comprehension when readers read multiple texts is relatively recent. It is suggested by certain researchers that in educational settings reading and obtaining information from multiple sources is an essential skill (i.e. Cerdan 2006). Anderson (1994 in Stahl et al 1996a) points out that in order to obtain a rich understanding of an event or concept, dealing with it through different perspectives is necessary and this can be achieved by constructing links across information presented in different texts, and this information and the links connecting different sources are remembered better. If the nature and quality of attainment and retention of information differ when readers read multiple texts, then the processes underpinning comprehension across texts may differ from comprehending one single text and such processes should be accounted for.

Among the reading models, one that moves beyond the conceptualisation of comprehension within the limits of a text (text base and situational model) is the '*documents model*'. In documents model, as well as the distinction between text base and situation model, the relations between texts and the integration of the situations the texts describe are also considered (Perfetti, 1997 and Perfetti et al, 1999). Perfetti et al. (1999) claim that readers can construct spatial and non-propositional representations that are not based on the explicit content of the text but as derivatives of it. Reading multiple texts produces representations that include connections of several kinds between the texts - implicit or explicit and based on complementary or contradictory information. The situation model from the first text is 'updated' when the information is confirmatory. When the information from the first texts is contradictory, the reader builds a new situation model that contradicts the situation model made on the basis of the first text. Therefore, the situation models from the two texts are marked as 'oppositional'. Moreover, document representations that are built on multiple, contradictory texts must incorporate information about the documents themselves, such as the document type, the author's identity, the date of publication, etc. The documents model suggests three levels of representation: *The intertext model* that links texts to each other through *intertext predicates* in terms of their rhetorical relations (support vs. oppose, agree vs. disagree, based on, relevant to, etc) and *the situations model* that represents situations described in texts with links between the texts (Perfetti, 1997, 346). When these two models are interconnected then we have a full *documents model*, a level at which texts are connected through a set of *document predicates* that specify the functional relations (as opposed to merely temporal and formal relations) among documents (e.g., Text A opposes Text B). Intertext predicates can be explicitly marked in a text, for example through citation. If one text is built on information learned through previous texts, it basically 'updates the situation model' created in the first text. The case is different when texts contradict each other implicitly: When the intertext predicates are not explicitly marked, the reader must infer an intertext model on additional knowledge of the texts. Implicit connections can be built through the situations connected to each text by the reader and this requires identification of document characteristics (Perfetti et al, 1995; Perfetti, 1997 and Perfetti et al, 1999). Documents are described by *document nodes* within the intertext model, which identify source characteristics, rhetorical goals and content (and subcategories of these) of the documents. Content at an intertext model is an abstraction of the main point or thesis of the document, a summary available to the reader as part of what he or she can come to know about a text. This may have become available to the reader as a result of building a situation model of the text. An intertext model will be connected to situations to form a full documents model in which texts and situations are situated in a coherent whole. Several factors such as reader characteristics and task factors determine whether a full intertext model will be formed or not.

Lacroix (1999) suggests that the comprehension of complex, multiple texts involves higher levels of cognitive processing that ensure the structuring of multiple text information as in hypertext reading. It is hypothesised that two distinct levels of macrostructural processing (forming of a condensed, coherent mental representation of the situation described in the text, see van Dijk and Kintsch, 1983) are at work when multiple texts have to be processed. *Macrostructure Construction*, as suggested by van Dijk and Kintsch (1978), involves identifying and hierarchising units of information through selection, generalisation and construction macrorules. Lacroix (ibid.) claims that this theory accounts well for the comprehension of a single text but may not be adequate to explain how mental representations are built from multiple texts. She suggests that *Macrostructural Organisation* is the process that accounts for the connection of several text representations through higher-level semantic links. Macrostructure Construction and Macrostructure Organisation are distinct processes that are affected differentially by specific textual and situational factors.

Stromso and Braten (2002) also comment that Kintsch's (1998) notion of each completed text becoming a potential component of the prior knowledge of the readers does not allow for sufficient specification of the sources that readers draw on when they try to understand multiple texts. Reading multiple texts is in their view '*discourse comprehension*'; a task of composing a new text by selecting, organising and connecting content from more than one source (see Spivey and King, 1994).

Several studies have also suggested strategies that seem to be peculiar to reading sessions when readers have to read and integrate information from multiple sources. Early surveys of the reading comprehension of multiple texts are grounded in the field of teaching history where, at the beginning of 1990s, the idea that learning from a single text could be an efficient way of learning about historical events was challenged. For example, Wineburg (1991) investigates how historical evidence from primary and secondary sources were evaluated by historians and high school seniors and identifies three main strategies the experienced historians used when evaluating information from multiple sources: *corroboration*, or comparing and contrasting documents with one another, *sourcing*, or looking first at the source of the document to evaluate, whether the bias of the source might have affected the content *contextualisation*, or situating the text in a temporal or spatial context to evaluate the extent the content of the document might have been affected by the time or place in which it was written.

Stahl et al (1996b) looked into the processes and outcomes of reading multiple original materials on a historical event and they found that not only the internal consistency of the students' mental models, but also the students' knowledge – knowledge of the relationships among concepts – grew as a result of reading multiple documents. Although the students put a good deal of reliance on the first text they read, they needed to 'select' useful information from among the irrelevant information to find what was important in the documents, 'reduce' the information to its gist, and, depending on the task, they 'produced evaluative-gist statements' which were considered to be conclusive statements that can be arrived at from reading more than one text.

Goldman (1997) and Goldman and Bloome (2004) also stress that in addition to searching for relevant information, multiple text processing involves 'other important skills such as note-taking, organising and coordinating information and detecting inconsistencies and redundancies across sources and otherwise integrating and evaluating information'. These literacy skills enable critical analysis of information; intertextual knowledge management, which 'entails understanding not only bits of information but relating these individual bits to one another in meaningful and systematic way (Goldman; 2004, 3-4). Goldman (2004, 337) found that the main difference in the text connecting activities of the students in her study was that when they were connecting information within the text they were currently reading, they were making causal connections but when connecting across texts, the students were comparing and contrasting information. Complex combination of information in multiple-text reading involved 'integration of multiple accounts with recognition of both commonalities of the basic events and the differences in perspective provided by each of the authors'. Goldman states that 'information across texts is a part of a larger whole not necessarily specified in any one of the texts' (ibid., 342-344).

On the other hand, Britt and Sommer (2004) state that learning by reading multiple texts involves the same processes and structures as single-text reading but requires additional skills for coordinating separate and sometimes discrepant presentations and for handling document level information. Since texts are not normally written to be read in conjunction with other texts, they lack explicit links to facilitate integration of information across texts. The demands on the reader to form a macrostructure are higher than when reading a single text with intratextual coherence. Britt and Sommer (ibid.) note that links that are not facilitated by the author of the individual texts may require additional resources and macrostructure formation of each individual text may be a prerequisite for successful formation of intertextual links.

There are also strong suggestions in the literature that building a complex, integrated understanding of multiple texts is associated with complex reasoning and deeper learning, and the degree of integration and learning is affected by task purpose (Braten and Samuelstuen, 2004; Cerdan, 2006; Gil et al., 2010; Goldman, 1997; Goldman and Bloome, 2004; Kobayashi, 2009; Perfetti et al., 1995; Stromso et al. 2008; Wiley and Voss, 1999) as well as by the readers' beliefs about what knowledge is, i.e. their personal epistemologies (Braten and Stromso, 2006, Braten et al., 2008, Stromso et al. 2008, Stromso et al. 2010). Moreover, integration of information across sources requires certain evaluative strategies that are unique to the documents model: Several recent studies (Braten et al, 2009; Braten and Stromso, 2010, Braten and Stromso, in press; Rouet, 2006) focus on the importance of 'source evaluation', evaluation of source characteristics when information from a text is processed in relation to other texts. Rouet (2006) claims that each document has a proper identity designated implicitly or explicitly by source information, and documents level reading includes – together with the representation of the situation within a text – information about how different texts relate to each other (Braten et al., 2009; Goldman, 2004; Perfetti et al 1999; Wiley et al., 2009). In reading across texts, as well as the similarities and differences between the content representation across texts, the qualities of the source documents are taken into account when the content is evaluated. The evaluation of the source document, that is assigning a role to each document as the descriptor of the situation, is determined as an important process in building a representation of a situation through reading multiple documents (Rouet 2006; 65). This is briefly referred to as '*source evaluation/ trustworthiness*' which involves the evaluation of such source characteristics as *content*, *document type*, *publisher*, *author*, *date* and *the reader's own opinion* (Braten et al., 2009).

Finally, documents model reading is considered to be socially constructed and likely to be shaped by the requirements of specific fields of study, each discipline transmitting its conventional processes of knowledge building within its own scientific environment (Hartman, 1995; Gil et al., 2010, Stromso et al., 2003).

Even in the brief review of the studies above, reading and integrating information across texts stands as one of the most complex cognitive tasks that a literate individual can perform. As it stands, intertextual reading has not been captured fully in any one reading theory, nor has it been possible to analyse it in its all possible aspects, let alone formulating an adequate theory for its assessment.

However, it is also not possible to neglect such a fundamental element of advanced learning, no matter what area of education is of concern. When the issue is preparing students for their learning in university settings or assessing their readiness for it, intertextual reading, as one of the most challenging language skills, must be given due attention. In the next section, two recent frameworks of academic reading will be discussed for their suitability as theories on which a construct of careful reading at sentence, text and multiple texts can be developed as a basis for its assessment.

2.4 Reading Models and Testing of Academic Reading in L2

As is stated above, any test development process should begin with recourse to a theory which accounts for the key aspects of the skill or ability in question. The selection of reading models and studies reviewed above (Section 2.3) deal with careful reading, putting the emphasis on different types of reading, such as sentence versus multiple texts, in an attempt to explicate the processes that take place at each level.

The existence of a variety of approaches to the explanation of careful reading is suggestive of the fact that careful reading may involve different cognitive processes at different levels. Therefore, in order to develop a valid reading test for academic purposes, we need a careful reading theory which is sufficiently capable of explaining the types of careful reading that have been discussed above: namely reading comprehension at sentence, text and intertextual levels. More importantly, for testing purposes, we should not only identify the cognitive processes underlying the skills but also contextual features that might shape the reading process, such as textual features. It is also of crucial importance to ensure that scoring procedures do not introduce any construct-irrelevant variables (Weir, 2005).

Recently, major EAP tests have introduced components that putatively tap the comprehension of multiple texts. However, frameworks which encapsulate the details of this type of EAP reading are scarce. The two reading frameworks that are now offered in the second language testing literature – Enright et al., 2000 and Khalifa and Weir, 2009 – do indeed make reference to types of careful reading involving different levels of processing. These frameworks suggest that different types of careful reading may require different cognitive processes according to the amount of text being processed. These will be outlined below.

Enright et al (2000, 5-6), proposing a framework of reading for the new TOEFLibt test, build on multi-componential conceptions of reading by adding the level of processing multiple texts. As stated in Section 2.2, their model encapsulates reading at four levels:

1) *reading to find information*: searching for specific information and comprehending discreet pieces of information. Rapid, automatic identification of words, working memory efficiencies, and fluency in reading are required.

2) *reading for basic comprehension*: understanding the general topic or main idea, important facts and details by forming some understanding of the main theme of the text without necessarily having to form an integrated understanding of the text. Reading for basic comprehension involves understanding a subset of individual ideas, some ability to construct a text

model representation of what is read and also the ability to form a relevant situation model.

3) *reading to learn*: understanding detailed information and connecting them into a coherent whole by understanding of cause-and-effect relationships, comparisons and contrasts, classification relationships, and persuasive intent. Reading to learn requires a more elaborated model of text construction, organising conceptual information, and suggests an efficient allignment of text model and situation model.

4) *reading to integrate information across multiple texts*: working across two or more texts and generating an organizing frame that is not explicitly stated. An intertext model of comprehension through generating a conceptual frame is required.

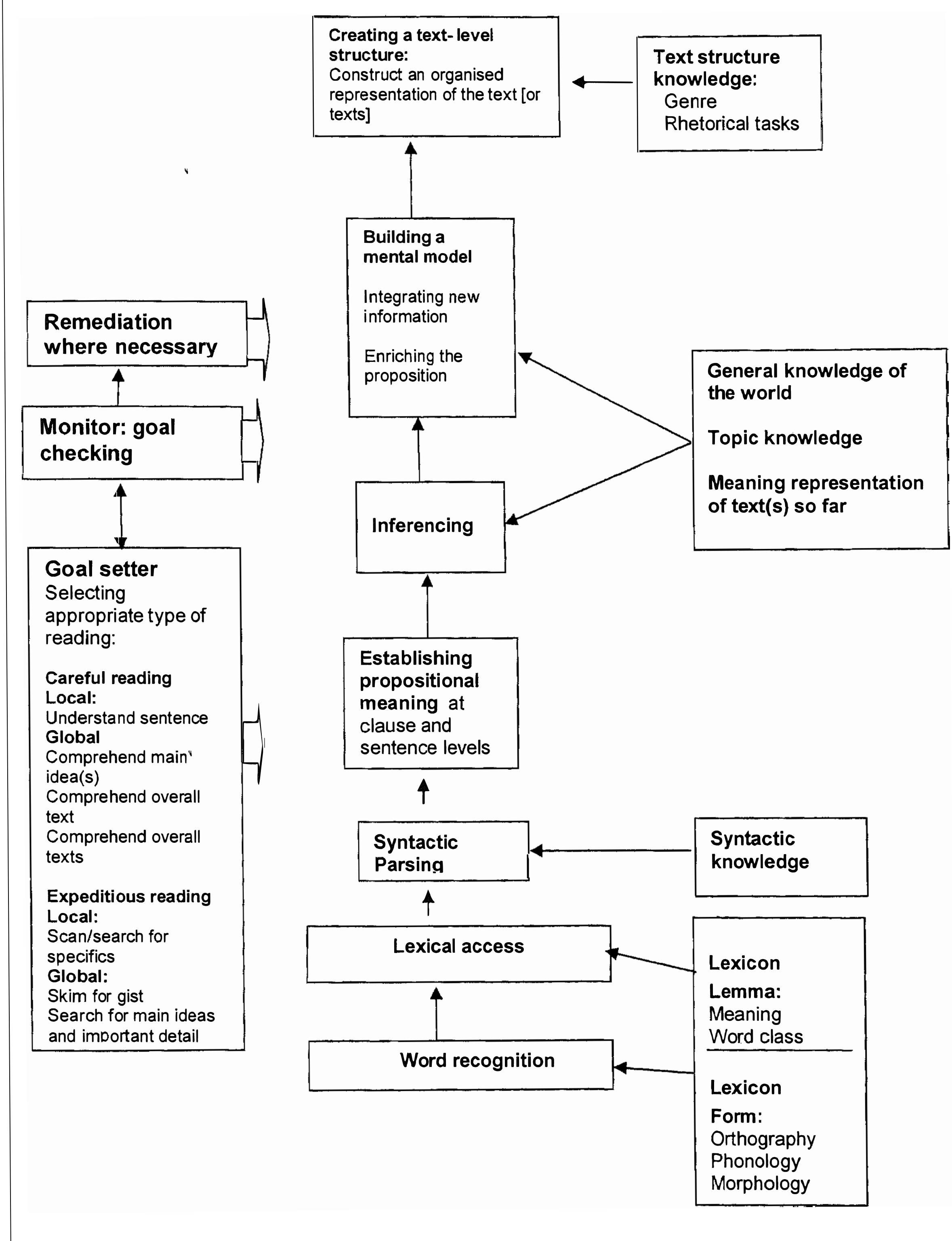
In Enright et al's (2000) model, reading for basic comprehension (2) is basically reading at the sentence or paragraph level (across sentences). Reading to learn (3) is the text level. The fourth level, as the name suggests, involves the processing of several texts; determining the relation of information between texts and integrating them in a written task. The levels of reading have been formulated from the 'reader's purpose' perspective as the guiding principle for test design, however Enright et al. (ibid., 2) claim that the processing perspective and the task perspective can also both be understood from, and inform the framework. The four skills are assumed to represent points on a difficulty continuum.

Khalifa and Weir (2009) capture the elements deemed important in the earlier frameworks, thereby accounting for the interactions between reader's purpose, cognitive processes and knowledge stored in long term memory. They hypothesise that difficulty in reading is a function of the level of processing required by reading purpose and the complexity of the text. Reading is conceptualised as having several types; expeditious versus careful and local versus global reading. Moreover, careful reading is further divided into four levels including within sentence (propositional meaning), across sentences (mental model; ongoing meaning making as the reader proceeds in the text), text (text model) and texts (documents model) models. Careful reading is a bottom-up process, starting with linguistic processing of the elements of a sentence and establishing propositional meaning (the literal interpretation of what is printed on the page). Through inferencing, the reader relates the message to the context. Inferencing is also functional in establishing coherence thus meaning between propositions as the reader integrates new information into a mental representation of the text so far. This is the stage where the reader starts to identify main ideas and impose a hierarchical structure on the information in the text. According to Kintsch and Van Dijk (1978) this is the stage where microstructure rules are at work to link the textual pieces and reduce the content to higher propositions to be stored in working memory.

Background knowledge on the content of the text and the meaning formed on the text so far facilitate inferencing and control of coherence and consistency in the text.

At the text level, micropropositions are collapsed into macropropositions. Recognition of the hierarchical structure of the text is of crucial importance in forming a unified understanding at the text level. The Khalifa and Weir (2009) framework acknowledges that reading at the intertextual level requires additional cognitive processes beyond what is required for understanding a single text, i.e. forming a global macrostructural organisation in which selected information is combined and structured through higher semantic links. In the framework, the monitor component is responsible for checking accuracy and appropriacy of comprehension and reader's reading purposes by making shifts in reading types when necessary. The knowledge base represents the several different types of background information that readers might need to incorporate in the reading process for successful comprehension (see Figure 2 for the framework).

Figure 2.2: The Khalifa and Weir (2009) reading framework



Both frameworks make reference to all three types of reading, though what is conceptualised in Enright et al. (2000) as 'basic comprehension' corresponds to Khalifa and Weir's (2009) 'building a mental model' rather than their 'sentence comprehension'. Khalifa and Weir (ibid.) explicitly distinguish sentence comprehension from 'the ongoing mental model', an intermediary level where sentence meaning enriched with inferences is combined with incoming information to develop the representation of the text. Therefore, unlike Enright et al. in Khalifa and Weir, the processing of information across sentences is distinguished from both processing information within a sentence and processing the text at the discourse level.

Both frameworks acknowledge similar reader purposes and cognitive processes for text and intertextual levels, however, both underline the need for more research for explicit designation of intertextual level cognitive processes. One advantage that Khalifa and Weir's framework has over Enright et al. is that it reflects the interplay between the readers' purpose in approaching the text and the cognitive level of reading they engage in. As a result, it better accounts for the adjustable, changeable and recyclic nature of reading comprehension through comprehension monitoring and goal evaluation components. As reading is a strategic process (Carver, 1997), these components are much needed in explaining reading behaviour under test circumstances as well. Moreover, the framework more clearly explicates sentence level comprehension and specifies the knowledge base that the reader needs to bring in at different levels of comprehension.

As it stands, Khalifa and Weir is a more comprehensive reading framework that accounts for the reading construct in more detail and therefore it lends itself more readily to the transfer of theoretical reading constructs to test specifications. However, more explicit accounts of the idiosyncratic cognitive aspects of intertextual reading need to be incorporated into the framework.

It should also be noted at this point that although the distinctions among the four levels of processing that Khalifa and Weir refer to in their framework are important in explaining reading comprehension, careful reading in this study is investigated in terms of the three distinct types identified in the cognitive models discussed above. The 'ongoing mental representation' (reading across sentences) and the 'situation model' (text) are treated together for the reason that to date, there is no adequate theory or research evidence that supports a distinction between the cognitive processes involved in an 'ongoing mental model' from those involved in a 'situation model'.

For the purpose of this study, it was more appropriate to combine reading 'across sentences' with reading at the 'text' level rather than combining the 'propositional' level and 'reading across sentences' because macroproposition formation starts at the ‘ongoing mental representation’ stage when reading across sentences just as it does in reading a text, but it does not occur in reading an isolated sentence. That is to say, when new information is integrated from a second sentence, the reader starts to build up an organised coherent whole that will be developed into a ‘text base’ which develops in an ongoing fashion throughout the reading process. In contrast, at the ‘sentence’ level, the reader interprets the input at the level of its local microstructure, analysing the structure and meaning of that individual proposition.

The distinction among several types of careful reading reflects the real life reading processes generally found in academic settings where readers find themselves having to read and learn from a whole text as well as integrating information from a variety of texts, especially in assignment preparation (Weir et al., 2009). It is obvious from the frameworks outlined above that careful reading as an umbrella term is seen as comprising processing at different levels: the propositional, situational and document models. As mentioned above, these constructs also appear in language tests with claims that students engage in these processes in their real life reading contexts and that construct valid reading tests should therefore tap into those different types of careful reading, too.

However, the distinctions between types of careful reading still need to be empirically validated and we need to establish whether processing a sentence, a text and multiple texts do in fact require different cognitive processes in assessment settings. Finally, we need to know whether there is a difficulty continuum as we move 'up' through these processing levels, i.e. in tests operationalising the three levels do we see a cline of difficulty upward from tasks requiring sentence level comprehension, through text comprehension to comprehension of multiple texts. The next chapter will offset out the methodology with which these questions are investigated in this study.

CHAPTER 3

METHODOLOGY

This chapter will give a detailed description of the methods of investigation and the rationale underlying the studies in relation to each research question. In relation to the first research question, it will inform the reader on the details of the questionnaire, reading diary and interview studies. The investigation of the second research question involved the development of a careful reading test, which necessitated several preliminary studies and these will be presented here in detail to explicate the test development process: document analysis - analysis of EAP tests, the development of context validity proforma for the investigation of textual features of academic tests, development of test specifications and development of test tasks. Besides, the methodological details of two main studies, test takers' verbal protocols and test data analysis, will be presented here. Verbal protocol study forms the pilot phase of test administration and it is also intended to gather evidence on the congruence between the test takers' actual reading processes and the test specifications. Detailed explanations will also be given concerning test administration and the analysis of test data. For the investigation of the third research question, the development of the cognitive validity proforma and the method of analysis of its data will be presented in detail.

3.1 Research Question 1) What are the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution?

Research Question 1 is formulated to explore the existence and use of careful reading at different levels in academic settings and to lay the grounds for further investigation into the different types of careful reading. The literature review has provided us with the theoretical background that careful reading can be analysed at different levels. In this section, the details of the questionnaire, reading diary studies and the interviews will be reported.

3.1.1 The Questionnaire Study

The review of literature presented in the previous chapter has made it clear that a high degree of efficiency and accuracy in reading is essential in academic contexts. Students are also expected to extend their knowledge through extensive reading by combining skills and strategies that are required for the different purposes of reading in tertiary level study. Therefore, it was important to have a detailed look at the reading activities of the students at the university level in order to understand whether actual reading behaviours of students confirmed these expectations.

An extensive study was designed to gain insight into the nature of academic reading for international students in the UK university context (Weir et al., 2009).[6] The researcher was responsible for the questions identifying careful reading activities and attendant contextual parameters shaping first year academic reading at the university level. The questionnaire in the study was intended to elicit information and views on the academic reading experiences of the students by identifying both the range of cognitive processes students employed when they performed the various reading activities for their university studies and certain performance conditions that shape their reading as well as difficulties they may face. Briefly, the focus of the questionnaire was on reading purposes, processes and difficulties.

The questionnaire was based on the reading framework proposed by Khalifa and Weir (2009). As detailed in Chapter 2, the framework identifies 'reader purpose', 'cognitive processes' and 'knowledge stored in the long-term memory' as the key elements of the reading process. Reading skills are conceived of on multiple dimensions as expeditious opposed to careful and local against global reading. Careful reading is divided into four levels including i) careful local within sentences, and ii)

[6] For the details of the piloting stage of the questionnaire, see Weir et al., 2009.

careful global across sentences (the mental model), iii) text (the text model) and iv) multiple texts (the documents model). The questionnaire reflected these reading processes and purposes of reading. There were also sections concerning personal details, important sources of information (articles, books, etc), whether the students read online or printed materials more and language difficulties they faced (see Appendix 3.1). After the pilot stage, the questionnaire was administered to 332 students in hard copy and to 434 students online, a total of 766 students over the period of five months at the University of Bedfordshire, UK.[7] The data were gathered from an opportunistic sample considered adequate to represent the population of students in the university: a total of 16150 students, including 6550 students in their first and 4400 in their second year. The sample included both home and overseas students, undergraduates and postgraduates and students in their second and first year of study at the university across a range of fields of study. Year 2 students were included as a check on whether reading processes altered much in subsequent university study. Unfortunately, the questionnaire could not be administered in more than one setting due to practical constraints and so longitudinal data were not available. However, for the immediate purpose, this was not seen as a problem since it should be possible to pin down different reading activities with these data to be able to lay the grounds for testing purposes. In relating the results in the next chapter, the responses to key variables are cross-tabulated according to

[7] The reader is referred to Weir et al., 2009 for the details of the pilot questionnaire study.

four groups: English as an additional language (EAL) and English as a first language (EL1) students, and Year 1 and Year 2 students and the results are given in percentages.

3.1.2 The Reading Diary Study

The diary study was intended to reflect in greater depth the choices students made in their use of reading types as they engaged in the process of preparing for a specific assignment. This provided a longitudinal, qualitative perspective based in specific study episodes to complement the extensive questionnaire data collected through the student survey although we acknowledge that the data came from a limited source

One home and five international students participated in the study and filled in diary forms (Appendix 3.2) over the course of a month while reading for their studies. The participants were all studying for a degree in Applied Linguistics and were preparing for their final dissertation; an extended research-based assignment. They were asked to fill in a diary form a few times a week in the course of their normal study routine. Each period of engagement with a text, whether brief or lengthy, counted as one reading session and was reported on one diary form. Between them, the six participants provided 62 diary forms, representing, according to the self-report data, approximately 1,500 pages of reading taking a total of 68 hours.

The diary form was developed by the researcher, revising the questionnaire that was based on the reading framework proposed by

Khalifa and Weir (2009). The framework was chosen, as discussed in Chapter 2, for its appropriacy for the analysis of academic reading. The reading purposes were detailed, cognitive processes presumably associated with reading processes were separated from strategies such as note-taking, and the wording was simplified. Overall, the diary form probed certain contextual parameters, reading purposes, cognitive processes, strategies to store information, perceived difficulties and usefulness of text choice (see Appendix 3.2).

In detail, *contextual parameters* refer to text source, quantity of text read, time spent, mode (on screen, on paper) and location (library, home). Section 2 concerns *purposes for reading*. It is assumed that the reader of a text may have different purposes for a reading activity ranging from *finding specific information* to *combining information across texts*. These purposes will influence the choice of reading type and hence the cognitive processes engaged (see Khalifa and Weir, 2009). The *cognitive processes* are addressed in Section 3. The dimensions of reading activity addressed include choices between different types (such as careful vs. expeditious reading) and different levels of processing (from the intra-sentential to the multiple text level). There is an assumption that certain reading purposes would most naturally trigger associated cognitive processes. For example, if a reader wants *to locate parts of the text that would be worth reading later*, he/she would presumably *look quickly for words relating to his/her assignment*. In this part of the study, the reading behaviours of six students reading for their dissertations were analysed and the links between the reading purposes and the cognitive processes were explored as they surfaced in the data. The responses in the diary form were reported in percentages and correlations between reading purposes and cognitive processes were analysed using Pearson's correlation.

3.1.3 The Interview Study

In order to get a closer view of the reading activities and the use of strategies of the participants in the diary study who were reading for assignment preparation, the six participants in the study were each interviewed once, after submitting their diary forms. The interviews provided an opportunity to follow up with the participants the issues that had emerged from the reading diaries and to carry out an additional, supervised reading session in which participants could comment in a concurrent and retrospective think-aloud procedures on the process of reading as they engaged in it.

The participants were asked to select an article in their subject area that they had not read before and to bring it to the interview session. On arrival, they were informed of the aim of the interview and were given a list of questions on which they were asked to comment as they read their text (see Appendix 3.3). The questions thus served to guide the participants' comments with the interviewer repeating the questions to focus attention as necessary during the interview. As the reading diaries appeared to have provided only limited information on how readings were chosen, the participants were also asked, in a follow-up activity, to use the world-wide web to find another article on the same topic and to judge its appropriateness for their purpose. The interview sessions were conducted individually and took approximately one hour. Each session was audio recorded and transcribed by the researcher. The data were then categorised according to reading purpose and associated cognitive processes following the categorisation in the diary form. Additional comments and strategies that were not captured by the diary form were also categorised and presented. The categorisation was repeated twice to ensure accuracy and to prevent oversights.

It is evident from the accounts in Rouet (2006) that manipulated situations where texts and tasks are designed to fit methodological purposes can hardly represent the large variations of reading in study contexts. The main aim here was to establish the levels of careful reading as they took place in a genuine academic setting. It was more important therefore to observe the reading operations used in response to an authentic academic writing task, rather than to manipulate the task for the purpose of extracting certain behaviours. In this study, through sampling behaviour from a reader's normal reading behaviour, a non-interfering observational methodology was employed.

3.2 Research Question 2) Do test takers score differently on tasks operationalising three careful reading types?

As mentioned above the probing of Research Question 2 involved in the first phase, a series of progressive studies that would lead into setting up the parameters for the design of the language tests, and in the second phase, the administration of the tests and the statistical analysis of the test data. The details of these phases will be given in order in the following sections.

3.2.1 Document Analysis

In order to answer whether test takers scored differently on tasks operationalising three careful reading types at academic level (careful reading at sentence, text and intertextual levels), firstly the major EAP tests available in the market were investigated to see whether there was any language test that covered the full range of reading operations that were needed for the purposes of this study. It is important to note at this point that although reading across documents is an activity heavily embedded in writing contexts, for the purposes of this study, which attempted to identify the levels of careful reading, an assessment tool that functioned without recourse to writing was essential. Therefore, the analysis focused on the assessment of reading skills at three levels, namely, reading at sentence, text and intertextual levels in the tests.

The document analysis consisted of a review of major EFL tests including TOEFLibt, IELTS, Cambridge ESOL Main Suite Exams (FCE, CAE, CPE), Pearson Test of English and Trinity College London ISE Test to evaluate the operationalisation of the careful reading construct in those tests. The manuals and test specifications were analysed and research studies that investigated the constructs in those tests were referred to when available (for TOEFLibt, IELTS, and Cambridge ESOL Main Suite Exams). The reported reading operations in the manuals or studies are given below for each test analysed.

.3.2.1.1 The Tests

TOEFLibt

TOEFLibt *The Official Guide* (2005) reports reading skills that are assessed in the test as below.

Reading for basic comprehension (sentence or across sentences): understanding vocabulary, pronouns, essential information in complex sentences, explicitly stated factual information within a sentence or across sentences.

Reading to learn (text level): recognising the organisation and purpose of the text, conceptualising and organising textual information into a mental framework, distinguishing major from minor points, and essential information from non-essential information, understanding idea relationships.

Inferencing questions (sentence to text level): understand an implicitly stated argument or idea, determine the purpose of the author and understand lexical, grammatical and logical links between successive sentences.

Reading to integrate information across multiple texts: This skill is assessed under the writing component of the exam as 'an integrated writing task' and the task requires understanding key ideas from a reading and listening text, determining relations between the two and accurately connecting information from the texts.

The first three levels of reading were analysed in a study by Cohen and Upton (2006) focusing on reading and test-taking strategies the examinees used in responding to TOEFL course preparation materials. Despite the carefully designed test specifications that putatively differentiate among the levels of careful reading, the associated items were not found to require and evaluate different academic skills but to assess similar components of academic reading at local level as well as test taking ability. The task that requires integration of information across texts has not been designed as a purely reading task and the integration of information is assessed through written performance. Therefore, TOEFLibt is not a test that assesses *reading* skill across multiple reading texts and as is attested by Cohen and Upton (2006), it is dubious that the reading to learn task assesses reading ability at text level either.

IELTS

For the analysis of the IELTS test, we refer to a detailed study, Weir et al. (2009), in which cognitive parameters of IELTS texts and tasks were analysed. Among the 1154 responses provided by two informed analysts, only 14 indicated that the test task could be completed by processing the whole text and no task in IELTS required multiple text comprehension. The rest of the items were based on comprehension at single sentence or across-sentences levels.

Cambridge ESOL Main Suite Exams

As analysed in Khalifa and Weir (2009), three Cambridge ESOL exams (FCE, CAE and CPE) cover such careful skills as:

FCE: reading across sentences

CAE: reading across sentences and text level

CPE: reading across sentences, text and multiple-text level

Among three tests, only CPE has multiple texts in the Use of English part. The task involves summarising two short texts on the same topic by selecting and linking information.

As it stands, CPE task is a documents level task that requires test takers to understand and relate the information in each text; however comprehension is assessed through an extended written output, which integrates 'writing' as a part of the construct measured in the test.

Pearson Test of English

Pearson Test of English has five levels, the highest three levels (3-5) putatively corresponding to Common European Framework of Reference (CEFR, Council of Europe, 2001) levels of B2-C2. Starting from the third level, the tests present multiple texts and integrated reading and writing tasks. Since there was not any manual explicitly reporting the reading operations assessed in the test at the time of the analysis, the researcher took the tests under exam conditions to report the reading processes and processing levels. The general distribution of tasks across the three levels in the manual (Edexcel, 2004) is determined as such:

Task Three a: comprehension questions on sentence and across sentences levels.

Task Three b: integrated reading-writing task based on two texts given in relation to the integrated task in which test takers are expected to locate, interpret and synthesise relevant information and transform it into an appropriate written form in response to the given writing prompt.

Task Four a and b (Levels 3 and 4): sentence or across sentence level comprehension questions on a third thematically related text.

Task Four b (Level 5): vocabulary at sentence level.

Task Four c (Levels 3 and 4): vocabulary at sentence level.

Task Four c (Level 5): pronoun reference at sentence level.

The first task, *Task Three a*, in versions 3 involves comparison of information between two texts; i.e. whether the information in the given statement appears in Text A or B or neither. In design, the task seemed to be geared at across-documents level and promising. However, upon completion, it became obvious that the task can be completed by searching the key words in each text and verifying the truth value of the information in the given statement by reading a single sentence. In *Task Three a* in version 4, the second question required the completion of an outline, a task which is generally claimed to test text level processing. However, as Cohen and Upton (2006) found for TOEFLibt, this task could be completed by search reading and sentence level careful reading. *Task Four* in the same version, a chart-filling task, can be processed at across-sentences level.

Task *Three b* requires the readers to read and select relevant information across two or three written texts (and optionally, one visual prompt graphics) to write an essay on a given topic. As it stands, this task is a legitimate documents level – discourse synthesis task mimicking an academic level essay task, however, the assessment is again done on an extended written output. The rest of the reading items are at local level, either sentence or across sentences levels.

Trinity College London ISE Test

The test has four levels (ISE 0-III) corresponding to CEFR (Council of Europe, 2001) levels of A2-C1 and all levels have a reading into writing task in which the candidates are given a text or texts to read and they are asked to complete a written task. The complexity of the texts and the writing task increases as the level increases. The test does not have a reading only component.

ISE 0: a reading into writing task with short and straightforward text.

ISE I: a reading into writing task testing the ability of the candidate to read and understand an authentic text and then respond to the information, the ideas and the opinions of the author.

ISE II: a reading into writing task in which the candidate is required to evaluate information and arguments from a number of different sources and develop an argument systematically. Candidates may be asked to synthesise information from a variety of text types.

ISE III: a reading into writing task in which candidates may be asked to synthesise information from a variety of complex text types. Information may be presented in different formats such as graphs, tables and diagrams. The tasks require the candidate to identify views and opinions which are not explicitly stated.

3.2.1.2 Final Remarks on Document Analysis

From the brief review of the tests above, it is evident that certain tests do not attempt to measure reading across texts (i.e. FCE, CAE, IELTS) and the ones that involve such a task do so through an integrated skills approach where the reading skill is tested through extended writing. In line with the conceptualisation of such integrated tasks, test takers are expected to compose from single and sometimes multiple documents by selecting, organising and integrating information. Granting that reading into writing is an important literacy skill to be measured in language tests and any inference we can make on the ability of the candidate to integrate and transform the information s/he has read into a written format is important, it should be remembered that in order to understand and assess the dynamics of reading at the level of multiple texts per se, we need a tool that assesses the skill on its own right. Therefore, at this point of the study, it became clear that such a test does not exist among available tests and needs to be designed from the very beginning.

It is possible to raise the question whether such a test may be redundant in the presence of more integrated reading-writing tests and to claim that intertextual reading processes can be measured through writing. However, performance in such integrated tasks is substantially dependent on the writing ability and considerable information may be lost in the cases where test takers written ability is low. On the other hand, in order to be able to conclude that a test taker has the necessary reading proficiency for university study on the basis of a reading test, it is evident that EAP reading tests should sample as many reading skills as possible, especially academically important ones.

3.2.2 Development of Context Validity Proforma and Text Analysis

The second investigation was related to the identification of contextual parameters. Most validation arguments (eg: Fulcher, 2003 and Weir, 2005) stress the importance of accurate description of the construct (reading operations in this case) we are attempting to measure and accurate description of the context (performance conditions) in which these operations take place. It is widely accepted that for valid inferences from test results, it is essential that target reading activities and test tasks be described in terms of both cognitive processes and contextual parameters (Weir, 2005, Weir et al. 2009). The description of the construct for this study depends largely on theory and has been established through the literature review and confirmed through the investigations into the real-life reading behaviours of university students (see Section 3.1). The description of the context particularly involves relating the features of test tasks to the language in the text that must be processed for the successful completion of the tasks. Thus, the features of texts are the key contextual features that should be described in detail. In the case of academic reading, this entails a thorough analysis and description of the target situation texts; university course books. The section (Section 3.2.2.1) below presents the research basis for the development of the context validity proforma and Section 3.2.2.2 explains the criteria for the selection of salient textual features to be included in it. Section 3.2.2.3 lays out the methodology for the analysis of university course book texts and Section 3.2.2.4 presents the analysis of the textual features in the texts. Comments on the analysis are given in Section 3.2.2.5.

3.2.2.1 The Research Basis for the Context Validity Proforma

In several validity frameworks, it is emphasised that evidence on the extent test tasks represent real world tasks is of crucial importance. Messick (1995a, 1995b) states that evidence on the content relevance and representativeness of assessment tasks can be gathered through specification of the knowledge, skills and other attributes revealed by the assessment tasks (specification of the boundaries of the construct domain to be tested). Bachman and Palmer (1996) argue that both the extent to which test tasks reflect important contextual features and the extent to which cognitive processes are similar to those used in the target language domain are important features of test tasks. Weir (2005) suggests that test tasks should be defined and approximated to real world tasks both in terms of contextual features and cognitive processes in order for the test performance to be generalised to target situation. Weir et al. (2009) emphasise that no matter how difficult it is to capture all the aspects of a target reading situation in an assessment task, features of texts used in an EAP reading test should reflect as many of the relevant characteristics of the target reading activities as possible since a lack of congruence between test and target situation texts may have serious implications on the interpretability of score outcomes.

Several studies focus on salient textual features that will impact on the comprehensibility of reading passages and test performance (see for example Alderson, 2000; Freedle and Kostin, 1993; Bachman et al., 1995; Fortus et al., 1998; Enright et al., 2000). Bachman et al.'s (1988, 1995) test comparison studies involve such textual properties as the *nature of text, length, vocabulary, grammar, cohesion, distribution of new information, type of information, topic of discourse, rhetorical organisation* and *illocutionary acts.*

Freedle and Kostin (1993, see also Freedle, 1997), in a detailed analysis of reading comprehension item difficulty, take into consideration *vocabulary, concreteness/abstractness, subject matter, coherence, length of various segments such as word, sentence, paragraphs* as text related variables. Fortus et al. (1998) investigated *length, number of negations, number of referential markers, vocabulary, grammatical complexity, abstractness, topic, rhetorical structure* as textual variables contributing to the level of difficulty of reading comprehension items.

Enright et al. (2000) identify two groups of salient textual features to operationalise in test texts: *grammatical/discourse features* and *pragmatic/rhetorical features*. Alderson et al. (2004) include *text source, authenticity, discourse type, domain, topic, nature of content, text length, vocabulary* and *grammar* as relevant features for text analysis. Khalifa and Weir (2009) suggest that linguistic demands of task input - reading texts in this case - can be explained in terms of *lexical* and *structural resources, discourse mode, functional resources, content knowledge* and *writer-reader relationships*.

Freedle and Kostin (1993) and Fortus et al. (1998) provide empirical evidence that a subset of the listed characteristics impact on the difficulty of reading comprehension tests. Moreover, recent advances in computational linguistics and the development of corpora have made automated analyses of textual features possible (see for example, Burstein, 2003; Crossley et al., 2007; Crossley and McNamara, 2008; Landauer et al., 2003)

In test development, readability formulas (including the Flesch Reading Ease and Flesch–Kincaid Grade Level, frequently used as estimates of textual complexity, have been critised as being crude measures of syntactic and lexical features such as word and sentence length. Such measures are seen as inadequate to reveal textual complexity (see, for example, Masi, 2002) and as being inappropriate for L2 readers (Brown, 1997). Masi (2002) suggests that together with such quantitative indicators of word and sentence complexity, other semantic and syntactic factors such as structural embedding, text type and the reader's content and background knowledge should also be taken into account in estimating the suitability of a text for a given reader, these latter characteristics being, of course, more difficult to measure. With the advancement in computational linguistics and corpora studies, several diverse analyses of textual features have been possible. Crossley et al. (2008) have proposed an alternative readability formula for L2 readers based on vocabulary frequency, similarity of syntax across sentences and referential cohesion. This formula proved successful at predicting scores on a suite of cloze tests.

Among the features analysed and stated as having impact on reading comprehension in the studies cited above, it has been possible for the purposes of this study to extract a list of quantifiable features that could be automatically analysed by using certain tools available on the Internet and another group of qualitative features that could be readily judged by test developers with minimal training. In this study, *context validity proforma* (see Appendix 3.4), the list of important textual features, was used in the analysis of target language domain texts, in this case, university first year course books and the findings formed the basis of contextual parameters in the test specifications on which the new test would be based. It has to be underlined that a practical, reliable and theory-based tool for the selection of test texts is a necessity in the field of language testing. The context validity proforma provided us the means to form solid test specifications in terms of contextual features and governed text selection process thereby securing content relevance and representativeness and authenticity in terms of contextual features

3.2.2.2 Identification of Salient Contextual Parameters

A more detailed analysis on the analytical comparison of the textual features of the university course book texts and IELTS reading texts was carried out in Weir et al. (2009) and Green et al. (2010), in which the researcher was principal researcher. The section below draws on those studies in discussing the relevant contextual parameters to be included in test specifications. Brief discussion on those textual features as factors impacting on reading comprehension is given below.

Text length

University students are expected to process long texts in relatively limited periods of time. Green et al. (2008) found that coping with this reading load under time pressure was a major cause of difficulty to students. Alderson (2000) and Nuttall (1996) argues that a long text is required for test takers to skim for main ideas, scan for specific information, make relevance judgements and distinguish between main points and minor details. However, as Alderson et al. (2004) point out in reference to the

CEFR (Council of Europe, 2001), distinctions between long and short texts are generally not clear. It is also not clear how long a text should be in order for test takers to employ such reading skills and what the time limits in relation to text length should be.[8] Obviously, the length of a test text is usually determined according to the purpose of the test and practical considerations such as other skills to be tested and the total duration of the test. The length has to be determined in test specifications. On the other hand, analysis of the length of the texts done under non-test conditions does not inform test parameters much.

Grammatical characteristics: Vocabulary and grammar

Vocabulary measures are said to be strong predictors of text difficulty (Read, 2000). However, simple measures such as word length do not inform us much about word complexity. Computer-assisted analysis of extensive language corpora has facilitated the use of word frequency lists such as Xue and Nation (1984) and Coxhead (2000) to inform language test development and validation.

The analysis of university course book texts in the present study with a web tool, *VocabProfiler* enabled investigation of several vocabulary features of the first year university texts. VocabProfiler provides a set of vocabulary measures that include *word length* (average number of

---

[8] In terms of careful reading, normal reading rate for L1 adult readers is accepted to be between 200-300 words per minute., i.e. 240 wpm in Just and Carpenter (1987). Crude as it is, this is usually taken into consideration in the timing of careful reading tests.

characters per word), *lexical density* (number of content words as a proportion of the number of grammatical words) and *word frequency levels* (the percentage of words occurring among the most frequent words in the British National Corpus; BNC) and *the percentage of words in a text also appearing on the Academic Word List* (AWL); sub-technical vocabulary. In the present analysis, *Standardized type-token ratio* (TTR: the ratio of types – or different words, to tokens – the total number of words occurring in the text) was also measured. TTR provides an indication of the number of different words the reader will need to know to understand a passage. The higher the TTR, the more diverse and complex the language of the text and therefore more demanding the text is. As TTR is affected by text length – the longer the text, the lower the TTRr– it is generally recommended that standardized lengths of text be used in calculation: WordSmith Tools were used to calculate TTRs based on 250-word sections of text.

Syntactic complexity is one of the major indicators of text complexity (Alderson, 2000). Texts with less complex grammar tend to be less difficult to read than texts with complex grammar (see Perfetti, 1997). Shiotsu and Weir (2007), using structural equation modelling, observed that syntactic knowledge played a greater role than lexical knowledge in accounting for variance in tests of L2 English reading. Therefore, a valid academic reading test should necessarily reflect the syntactic features of the texts in the target situation.

Syntactic complexity has been either judged by readability formulas that depend on sentence length or through subjective expert judgement. Alderson et al. (2006) provide four categories of grammatical complexity in distinguishing between texts suited to different proficiency levels (from only simple sentences to many complex sentences) and Khalifa and Weir (2009) have used these in the analyses of the Cambridge ESOL General English Examinations. In the present study, syntactic complexity is measured through quantitative analyses. Indices of rough measures such as Flesh Kincaid readability formulas and more sophisticated measures from Coh-Metrix are used (Graesser et al., 2004; McNamara et al., 2005).

Readability statistics (*Flesch Reading Ease* and *Flesch-Kincaid Grade Level*), widely used in test development, are available through Microsoft Word: both measures being based on the relative numbers of syllables, words and sentences found in a text. Flesch Reading Ease scores range from 0 to 100 with lower scores reflecting more challenging texts. A score below 50 is said to require college-level reading skills. The Flesch-Kincaid Grade Level is based on the US school system, with 12 representing the final year of High School and 13 to 16 the college level. The Crossley, Greenfield and McNamara (2008) readability formula mentioned above as a potential alternative to traditional readability measures for L2 readers is also included here as *Coh-Metrix readability*. Coh-Metrix readability integrates *lexical* (logarithmic frequency of all content words in CELEX corpus, Coh-Metrix index 41), *syntactic* (syntactic similarity of adjacent sentences, Coh-Metrix index 24) and *meaning construction* (proportion of content words that overlap between adjacent sentences, Coh-Metrix index 13) *indices* to yield a more reliable and meaningful readability score (Crossley et al., 2008).

Cohesion and rhetorical organization

Although the effect of the use of cohesive devices on comprehension is less clear-cut than for grammar and vocabulary, there is evidence that explicit cohesive devices help in establishing textual coherence (Goldman & Rakestraw, 2000), and their absence inhibits the recall of texts, and this is indicative of a less successful mental representation (Ehrlich, 1991). Barnett (1989) suggests that rhetorical features must be integrated in the analyses of text difficulty and studies investigating the effects of textual organization on text recall (see, for example, Carrell, 1984; Goh, 1990) suggest that differences in rhetorical organization affect comprehension. In these studies, certain rhetorical patterns such as problem-solution, comparison, and causation structures were found to result in better recall than classification or description structures. Koda (2005) cites a number of studies reporting the positive effects of improving text structure and Freedle (1997) finds that texts subjectively judged to be high in coherence yield easier main idea reading comprehension items. In this study, coherence is assessed quantitatively through two Coh-Metrix indices;*content word overlap*, the proportion of content words in adjacent sentences that share common content words (index 13), and *LSA mean all sentence similarity* (index 15). LSA index is computed based on the semantic and conceptual similarity among all the sentences in a text, the higher the score, the higher the similarity between parts of the texts, therefore, the higher the coherence is.[9]

Rhetorical Organization refers to the extent to which there is an explicit pattern of topic progression through the text. Such progression might be signalled by headings, topic sentences and discourse markers. The *explicitness of rhetorical organisation* is judged qualitatively in the present study and is intended to reflect the ease or difficulty with which the overall propositional pattern of the text might be understood by the reader.

Genre and rhetorical task

Genre is explained by Weigle (2002, p. 62) as the expected form and communicative function of the written product. Genre is generally understood to encompass 'salient features and conventions which are shaped by communicative purposes' (Hyland, 2000, p. 62). Therefore, specific genres will involve specific conventional features (lexico-grammatical, semantic, and discoursal) which are likely to impinge on the text processing of readers (Bhatia, 1997; Hyland, 2000). For this reason, it is a usual practice to select texts for EAP reading tests from university

[9] Indices 13 and 15 are chosen for the analysis among the other cohesion indices provided by Coh-Metrix as index 15 was the only cohesion index that indicated a significant difference between two corpus of texts in Green et al. (2010), and Crossley et al. (2007) found that content word noun overlap, index 13, was a strong predictor of text adaptation for L2 learning.

course books so that lexical, syntactic and discourse features specific to course book genre will be reflected in language tests, too. Enright et al. (2000) list main genres relevant to language testing *as text book*, *magazine/newspaper article*, *research/academic journal article* and *report*.

Classification of *rhetorical task* was based on Enright et al.'s (2000) classification of 'pragmatic features'. Rhetorical task basically referred to 'the primary intent of the author' that guides the reader in understanding the text (Enright et al*.,* 2000, p. 20). Enright et al. suggest a three-way classification: *Exposition* informs the reader. It may involve descriptions, comparisons, contrasts, explanations and elaborations. *Argumentation/persuasion/evaluation* supports a point of view with reasons, evidence and analysis of an opponent's errors in reasoning. Vocabulary might reflect attitude or perspective and it may be personal in tone. An argumentation text departs from a balanced, unbiased stance. *Historical biographical/ autobiographical narrative* tells a story with a defined setting and episodes. Evaluation of rhetorical tasks are done qualitatively in this study.

Subject area, subject knowledge and cultural knowledge

The impact of a reader's knowledge of the topic of a text on his or her comprehension is widely acknowledged (Nuttall, 1996; Khalifa, 1997). Enright et al. (2000) suggest that test tasks should be based on materials sourced from a variety of subject areas so that effects of topic familiarity can be minimised. Studies such as Chihara et al. (1989) and Sasaki (2000) have provided evidence that cultural knowledge plays an important role in L2 text comprehension. Although academic text books will normally be expected to contain knowledge of specific discipline areas and of particular cultures, it is an accepted practice to avoid culture or discipline specific content in language tests (Alderson, 2000). Words appearing in the classes of less frequent, technical or off-list words, place and brand names can be identified in vocabulary analysis mentioned above. Otherwise, judgement on subject specificity and cultural knowledge has to be gathered qualitatively.

Text abstractness

Alderson et al. (2006, p. 127) see the degree of abstractness as a useful feature to consider in estimating text difficulty in relation to the CEFR (Council of Europe, 2001). Information that is more abstract may prove to be more difficult to process and often implies a linguistic complexity that may further cause difficulties for readers. Moore and Morton (1999) make the observation that, with variation across disciplines, much academic text is concerned with the theoretical treatment of abstract phenomena. To reflect this, a proportion of texts included on tests of academic language ability would be expected to concern abstract ideas. In this analysis, text abstractness was quantitatively measured by one Coh-Metrix index, *mean concreteness of content words* (index 44) based on Coltheart (1981).[10]

[10] Coh-Metrix index, *mean concreteness of content words* (index 44) was again the only index that showed the difference between the two sets of texts in Green et al. (2010).

The concreteness value ranges between 100-700, the higher the value, the more concrete the words in the text.

3.2.2.3 Text Analysis

Using the context validity proforma developed with the inclusion of the parameters discussed above, textual features of first university course books were analysed. In order to perform the analysis, 14 core first year undergraduate textbooks at the University of Bedfordshire were selected to represent key academic texts that incoming international students would need to be able to follow. These were the books that lecturers and students reported to be essential reading, had most often been taken out from the library in the current academic year and had the highest number of library reservations over the previous three years (Weir et al., 2009). Three extracts were taken from each of the 14 textbooks to provide 42 samples of academic text. These sections were self contained passages that could be understood as coherent stand alone texts and were extracted from the opening chapter, the middle chapter and the concluding chapter of each book. The length of the extracts changed between 500 and 1500 words.

As stated above, the texts were analysed through automatic tools where possible and in line with Bachman et al. (1995) and Alderson et al. (2006), through expert judgement when they were not measurable automatically.

The two experts, one this researcher, had PhDs in applied linguistics and experience of teaching and test development in the area of academic reading. They discussed the criteria and piloted the cognitive validity proforma on a set of five texts taken from other sources. These judges then used the Likert scales and classification tools to evaluate the texts.

3.2.2.4 Analysis of First Year University Course Books

The analysis of text books yielded certain text characteristics observable in university course books. The results of the qualitative evaluation of the texts are given in percentages, and the parameters calculated through automatic tools are given in Table 3.3

The mean values of responses to *rhetorical organisation*, *subject specificity* and *cultural specificity* items are given in Table 3.1 and rates of agreement between the two experts on those items and *rhetorical task* item are shown in Table 3.2. Though the exact agreement percentages (31-52%) on these parameters were not high, when percentages were calculated allowing plus or minus one point discrepancy, the judges agreed on the issues of rhetorical organisation, subject specificity, cultural specificity with 93-87%. The expert judgements on *genre* and *subject area* parameters were not included in this analysis as the texts were all from *text books* and they were chosen from a variety of disciplines. In making classification judgement on *rhetorical task*, they were in 80% agreement.

Table 3.3 shows that in terms of rhetorical organisation, the university course book texts had quite an explicit structure (2.21/5), however culturally somehow specific (3.18/5) and quite specific in terms of subject matter (3.69/5). Both experts agreed that most of the texts were expository in nature – 27 out of 42 undergraduate texts and they agreed that argumentation and historical/biographical texts were also represented among the set of texts.

Table 3.1eMean values of the responses by experts on cognitive validity proforma

Parameter	Meand 5	Std. Dev.
Rhetorical organisation	2.214	0.54
Subject specificity	3.690	0.94
Cultural specificity	3.179	1.08

Table 3.2: Rates of agreement between the two judges on text characteristics

Parameter	Exact	+/- 1
Rhetorical organisation	52%	93%
Subject specificity	31ø%	87%
Cultural specificity	33%	89%

The figures in Table 3.3 are mean values of the textual features identified in context validity proforma and yielded by Wordsmith, VocabProfiler and Coh-Metrix tools. In the Table, we see that the average word length is 5.14 characters, STTR, 51s64% and Lexical density is 0.56. 74% of the words in the texts come from the most frequent 1000 words in BNC and approximately 12% come from the the most frequent 2000 words. There are fewer words from the class of 3000 words. However, 4.33% of the words in the corpus are among the least frequent class of 15000 words. 10.51% of the words are among the subtechnical vocabulary list. Classical readability scores (AWS, ASP, FRE, FKGL) show that the texts are located at the college-end of the specturum (FKGL: 12-16), yet the average difficulty is not towards the very end. For the concreteness value of the words in the text, we can say that it is at the mid point of 100-700 scale. However, as there are no determined values for *Coh-Metrix readability*, *content word overlap* and *LSA, sentences all combinations mean*, we cannot yet judge what these values may mean; i.e., whether the texts are strongly coherent or not. Despite this, we can say that these are the values the analysis of 42 university course book excerpts yielded and they may form guidelines in setting up the test specifications in terms of contextual parameters and in selecting the texts for EAP tests.

Table 3.3: Quantitative analysis of texts

Parameter	Mean	Std. Dev.
Average character per word (ACW)	5.14	0.36
Standardised type token ratio (STTR)	51.64	4.13
Lexical density (LD)	0.56	0.05
1000 word frequency BNC %	74.00	6.59
2000 word frequency BNC %	11.89	3.16
3000 word frequency BNC %	2.62	1.10
Frequency< 15 K	4.33	3.18
AWL level	10.51	3.47
Average words per sentence (AWS)	21.47	4.26
Average sentences per paragraph (ASP)	3.35	1.14
Flesh Reading Ease score (0-100) (FRE)	36.82	13.08
Flesh-Kincaid Grade level (FKGL)	13.66	2.34
Coh-Metrix readability	12.71	3.54
Proportion of content words that overlap between adjacent sentences	0.10	0.03
LSA, sentences all combinations mean	0.26	0.07
Concreteness, mean for content words	357.62	21.74

3.2.2.5 Final Remarks on Context Validity Proforma

In general, it is obvious that an attempt to quantify text analysis and relating it to context (i.e. content) validation is a worthwhile process as it reduces the risks that subjective human judgement may bring in and it makes the procedures replicable thus more reliable. Both in the field of language assessment and teaching there is a strong need to quantify judgements and context validity proforma stands as a research-based practical tool. However, as computational analysis of texts is a very new field and the use of it in determining the textual features for EAP tests is a unique attempt, there is no accumulated information with which we can verify the findings in this study. The selection of certain parameters among a multitude of options, especially Coh-Metrix indices, had to depend on scarce literature available and subjective criteria. Therefore, we do not know yet whether the indices used in this study cover all the necessary features of academic texts and whether all can be meaningfully used in EAP testing. Obviously, there is a strong need for large scale research that will investigate several aspects of the issue such as how textual features relate to, for example, test performance. Nevertheless, as it stands, context validity proforma proved to be a very useful tool that formed the link between the theory and practice by ensuring theory-based analysis of target situation texts and the adaptation of the features of those texts to testing situations. This is substantial evidence for content relevance and representativeness.

Another issue is that in the case of language testing, it is questionable whether all the properties of target situation context should be represented in tests. While there is authenticity on one hand, there are the issues of irrelevant specificity on the other. It is a general contention in the field of language testing that language learners cannot be held responsible for subject specific content or culture specific information, and texts with such elements pose additional and unwanted difficulties for the readers. Therefore, reading texts should not assume too much subject specific and culture specific background from test takers. In a parallel study comparing the textual features of university course books and IELTS texts, Green et al. (2010) have found significant differences between the test texts and course book texts in several aspects; higher rates of subtechnical vocabulary, higher rates of vocabulary from the least frequent classes, and higher subject and cultural specificty and higher abstractness on the part of the course book texts. They conclude that the differences between IELTS and course book texts reflects the requirement on IELTS to avoid technical and cultural allusion and less abstract and theoretical approach to topics in it (p. 206). Therefore, before all the textual features can be taken for granted for EAP test texts and transferred to test specifications automatically, they should be evaluated against assessment criteria. Green et al. (2010, 204) rightly suggest that 'final decisions on the suitability of material will inevitably require human judgement and test developers will need to weigh the impact of text features in relation to the purpose of the test.' In Section 3.2.3, where the development of test specifications is discussed, these findings will be re-evaluated in terms of their appropriacy for the purpose of the test to be developed.

3.2.3 Development of Test Specifications

The third stage in the investigation of the second research question was the development of test specifications. Development of test specifications is an intermediary but a crucial step in test development as the context validity – content aspect of construct validity in Messick's (1989) terms – is ensured by the evidence of content relevance and representativeness whereby cognitive (construct) theory specifies the boundaries and facets of the reading processes and test specifications build up the implicit links between cognitive processes measured by the test and the theory.

Since test specifications based on the theory will guide test construction, the relevance and representativeness of the test tasks are closely linked with the relevance and representativeness of the test specifications. Therefore, the construct theory delimits the domain specifications, which in turn determine test specifications. As mentioned above, test specifications are important in determining item or task relevance and representativeness. It should also be remembered at this point that for Messick (1989, 34), two major threats to construct validity are construct-underrepresentation and construct-irrelevant test variance. Construct-underrepresentation occurs when the test is too narrow and fails to include important dimensions or facets of the construct. Construct-irrelevant test variance, on the other hand, stems from construct-irrelevant difficulty or construct-irrelevant easiness. The former is taken care of by ensuring that the breadth of specifications for a test reflects the breadth of the construct invoked in score interpretation (i.e. test specifications are adequately based on the construct theory). The latter is usually investigated by item performance evidence (item analysis).

The function of test specifications in warranting the validity of a test should be clear in the discussion above. Construct theory that a test will be based on should represent the domain clearly and accurately for any test based on it to be useful. Grabe (2000) stresses that any comprehensive theory of reading should account for linguistic, processing, learning, social, affective and motivational aspects of reading. However, as discussed in Chapter 2, although it is one of the most researched areas with various aspects analysed, it has not been possible to encompass all the aspects of reading in one theory, nor is there a broad enough theory of assessing reading.

However, Khalifa and Weir (2009) as extensively discussed in Chapter 2, consolidate cognitive, linguistic and contextual aspects of reading into a more unified framework of reading assessment. In the framework, the place and role of linguistic elements at different textual levels (global versus local) along with different reading skills and strategies (expeditious versus careful) are investigated and explicit references to contextual variables are made. The framework accounts for such important variables as reader purpose, comprehension focus, text coverage, rate of reading and relationship with underlying processes and it combines premises of Kintsch and van Dijk (1978) and Just and Carpenter (1987) to adequately account for the reading behaviour both at the local and global levels using expeditious and careful reading processes. The goal setter element accounts for the metacognitive activities in deciding the level of processing that the text will be approached by the reader, and the components of the knowledge base explain the elements required for comprehension. In terms of careful reading, careful reading processes at the local and textual levels are explicitly accounted for and careful reading at intertextual level has been acknowledged as an additional level of processing, however with less detail than for the other two skills.

With careful reading at the local and textual level, transference of skill definitions to test specifications is apparent. Especially with the elaboration of the cognitive component into three dimensions (goal setter, processing levels and knowledge base) and the unification of contextual parameters into the reading process, the framework has become more dynamic and comprehensive in comparison to its Urquhart and Weir (1998) version. Based on the present version, it is possible to determine the reading purposes and processes, the amount of text that need to be covered and the features of the text to be processed, all of which have direct relevance for test specifications.

However, the processes at the intertextual level are less explicit. It is not clear in the designation of the textual and intertextual levels whether the comprehension processes are similar. For example, does the use of background knowledge involve similar processes at both levels? Are links connecting parts of a text and links connecting different sources the same? When a text is processed incrementally, elements of the texts are put in a hierarchical relation with each other. Are the elements of a subsequently read text accommodated in the same hierarchical template? When the information in two distinct texts are contradictory, are the same linking processes as in a single text used, or are there any peculiar evaluative strategies for linking contradictory information from multiple texts? Or, does each completed text become a potential component of the reader's background knowledge as it is claimed by Kintsch (1998)?

In order to form test specifications for intertextual level, some, if not all of these issues should be determined. However, not only in Khalifa and Weir (2009) but also in the existing literature, there are few explicit data about the precise nature of reading across documents. Therefore, we had to refer to the available research specifically focusing on the documents model and integrate findings from the studies conducted to investigate the nature of academic reading above (Section 3.1) to the extent that they inform test specifications.

In sum, the Khalifa and Weir (2009) framework forms a pertinent and applicable theoretical starting point for the description of the behavioural domain and thus for test specifications. In line with this framework, the test specifications for this study included descriptions of reading purposes, related cognitive processes, processing levels and contextual features. At the intertextual level, suggestions from the literature and the present research were incorporated where necessary. This was done by the researcher and a professor in language assessment in an iterative manner in which the findings from the questionnaire, reading diary and interview studies were re-evaluated in terms of cognitive processes and processing levels and compared to the premises of Khalifa and Weir (2009) framework. Hence, reading operations and the processing levels as well as their use in academic reading were confirmed.

Secondly, the reading descriptors in Common European Framework of Reference (Council of Europe, 2001a) were analysed and the reading descriptors at B2 and C1 levels were analysed to cross-validate the reading operations and processing levels identified for this study as most higher education institutions in the U.K. accept B2 or C1 level of English proficiency as sufficient for entry. The reading operations in the test specifications were framed taking into consideration reading descriptors at C1 level. Contextual features in the test specifications were drawn from the analysis of undergraduate course books analysis presented above (see Section 3.2.2).

Section 3.2.3.1 below presents the test specifications for careful reading at sentence, text and intertextual levels. Section 3.2.3.2 gives the final refinements on contextual parameters to be established in test specifications.

3.2.3.1 Test Specifications for Careful Reading at Sentence, Text and Intertextual Levels

The test specifications for careful reading at sentence and text level come directly from the model:

Careful reading at sentence level

Sentence comprehension is explained with word recognition, lexical access, and syntactic parsing and these are reflected in test specifications:

Reading purpose: to comprehend meaning formed with in a sentence; understanding words, pronouns, and syntactic units within a sentence.

Cognitive processes: word recognition (accessing orthography, phonology and morphology of the words), lexical access (accessing the meaning of the words in a mental dictionary), syntactic parsing (analysing phrases and clauses at syntactic level), establishing meaning (forming a proposition).

Processing level: Local, sentence level

Careful reading at text level

Text comprehension involves accurate processing of the hierarchical relations of the elements of a text – relations of sentential elements and relations between sentences as well as relations between groups of sentences, i.e. paragraph structures and functions – in arriving at an organised and coherent mental representation of the whole text. The readers are assumed to resort to their background knowledge of several types - lexical, syntactic knowledge as well as prior information at higher levels on content and text structure. From this description, the following test specifications were drawn:

Reading purpose: to comprehend overall information content of a text, to form a unified understanding (a mental summary) of a text

Cognitive processes: proposition formation, organisation of propositions, establishing meaning between propositions by integrating new information, building a mental model (macrostructure at discourse level), building a situational model that incorporates knowledge base.

Processing level: Global, text level

Careful reading at intertextual level

From the literature discussed in the previous chapter, several suggestions relating to the purpose of and the cognitive processes involved in intertextual reading can be gathered: Reading multiple texts requires building of connections of several kinds between the texts. These connections between the texts can be explicit or implicit and the information from the texts can be complementary or contradictory. Therefore, the connections between the texts are not only temporal and formal but also functional. Reading across texts also requires incorporation of information about the documents themselves, such as document type, the author's identity and the date of publication. In multiple text reading, texts are compared and contrasted with one another, sources are critically evaluated, and situated in a temporal or spatial context to evaluate the extent the content of the document might have been affected by the time or place in which it was written (Wineburg, 1991). Readers need to 'select' useful information among irrelevant information to find what is important in the documents, 'reduce' information to the gist, and depending on the task, 'produced evaluative-gist statements' which are considered to be conclusive statements that can be reached from reading more than one text (Stahl et al. (1996b). Combining information from different texts involves different processes from macrostructure formation on a single text since the separate texts do not normally include explicit links between them to facilitate integration of information across texts (Britt and Sommer, 2004).

Reading at intertextual level, as manifest in the above summary, requires a multi-faceted and complex process of information manipulation. As it is next to impossible to reflect all the facets of intertextual reading in a single reading test, the definitions of the process in several documents were focused on to elicit the core purposes and processes unique to reading across texts. The following were extracted:

* integrating informational content of a text into the wider context of the topic (Stromso and Braten, 2002)

* synthesising information: integrating information from multiple texts by selecting, organising (supplying a new organisational structure) and connecting information (by providing new links between related ideas)

from multiple texts and synthesising information by producing a new text (Spivey and King, 1987)

* forming a representation of the situation within a text and forming new types of connections between different texts through evaluating how they relate to each other (by understanding agreement, disagreement, discrepancy and other such qualities between the content representations across texts) and forming a new mental representation of the situation (Braten and Stromso, 2003; Goldman, 2004, Perfetti et al., 1999)

* understanding relations between texts by comparing and contrasting information and actively building new semantic links by evaluating relative standing of information against each other, creating an entirely new situation beyond what was described in the texts, organising and situating incoming information in the mental model of a new text which combines situations across texts and itself (present study)

Table 4.17 in Section 4.1.4 of the study summarises the findings from the interview study (see Section 3.1.3.), one of the studies integrated in the test specification development process.

Table 4.12: Summary of Reading for Writing Purposes

Reading Operation	Purpose	Process	Outcome
Expeditious Reading	Finding relevant information	Word, meaning matches, textual organisation clues Background knowledge on content and textual organisation	Selection of information for careful reading
Careful Reading at Text Level	Understanding the argument	Microproposition formation Textual semantic links Macrostructure formation Background knowledge of all types	Situation model formation
Careful Reading at Documents Level	Understanding and integrating multiple arguments raised on the same issue	a) Background knowledge from previous text Intertextual semantic links between texts b)Documents knowledge (Source characteristics)	Situations Model + Intertext Model = Documents model formation
Evaluation of Information	Inclusion in the task	Intertextual semantic links between the texts and the task Judgements of usability	Formation of a unique situation model

The definitions were still suggestive of a multitude of cognitive operations clearly due to the complex nature of reading to integrate information from several sources, which is usually done to produce a written text in the end. By definition, some of the operations are related to the 'reading' phase of the process and some, to the 'writing' phase. What is of interest to us are the processes that are within the confines of comprehension and integration of information rather than the use of it in a written product. Therefore, as cognitive processes that are related to intertextual level per se, the phrases below have been filtered through the definitions and repeated below:

* connecting information (by providing new links between related ideas) from multiple texts

* forming new types of connections between different texts through evaluating how they relate to each other (by understanding agreement, disagreement, discrepancy and other such qualities between the content representations across texts) and forming a new mental representation of the situation

* understanding relations between texts by comparing and contrasting information and actively building new semantic links by evaluating relative standing of information against each other, creating an entirely new situation beyond what was described in the texts.

What is common in all these definitions is that the reader has to detect and understand the semantic links between the sources and has to identify the nature of those links. Two distinct pieces of information are then integrated into a new, updated situation. The reader is then involved in a writing-geared evaluative process which is shaped by task requirements (written assignment). Braten, et al. (2009), and a few other studies (including the present one), suggest that in the creation of the new mental model, source characteristics play an important role. However, the successful use of source characteristics is considerably affected by the expertise of the reader and the firmness of his or her background knowledge in the field, hence are not solely dependent on the comprehension ability of the reader. Therefore, the cognitive processes relevant to intertextual reading are stated as 'detecting and understanding the semantic links between the sources and identifying the nature of those links, integrating pieces of information into a new mental model' in this study:

Reading purpose: integrating informational content of a text into the wider context of the topic,

Cognitive processes: detecting and understanding the semantic links between the sources and identifying the nature of those links, integrating pieces of information into a new mental model,

Processing level: Global, intertextual.

3.2.3.2 Contextual Parameters

The analysis of undergraduate course books described in Section 3.2.2.4 has yielded several textual features that formed the contextual parameters in the test specifications for the tests to be designed for this study. However, as mentioned in Section 3.2.2.5, not all characteristics of undergraduate texts can readily be transferred to test tasks. The test texts should not be too specific in terms of subject and cultural elements, which otherwise would pose an unnecessary difficulty for readers. Although they may assume some academic content therefore academic vocabulary, cultural elements should be reduced to a minimum. Cultural and subject specificity is usually reflected in the vocabulary items in texts. In the VocabProfiler analysis of word frequency classes, the words in the least frequent classes and off-list words will yield such features. Comparing these values with the ones in Green *et al.* (2010), it was decided that in the test, AWL words could account for up to 10% of a text, but that very low frequency words (i.e. off-list words) should remain at 1% at the most. Otherwise, other textual characteristics were taken as they appeared in the university course book analysis in Section 3.2.2.4. See Appendix 3.5 for complete test specifications.

3.2.4 Development of the Test Tasks

To establish more clearly whether there were differences in both process and product in careful reading activities at three levels; i.e., sentence, text and intertextual levels, three versions of an academic reading test each comprising three test tasks (Task S: sentence level, Task T: text level, Task I: intertextual level) were designed. It was hypothesised that each task type would make differing demands on students which would be reflected in different levels of performance and different patterns of response and different reported reading types on the part of test takers. Section 3.2.4.1 will explain the selection of test texts for that purpose, and Section 3.2.4.2 will present in detail the design of the test tasks.

3.2.4.1 Text Selection

Each task in the test was designed on a separate text to be able to measure the task performance without contamination from any unexpected reading behaviour that might surface as test takers increased their familiarity with the text. Cohen and Upton (2006) report that reading strategies and skills test takers use in attempting tasks may change in nature as test takers become more familiar with the text and use, for example, more search reading strategies rather than reading carefully. A second consideration was a possible text effect. To try and control text effect, each task was matched with a text yielding a three-version set. This showed whether or not it was the different texts that affected different performance or the tasks.

The texts to be used in the study were selected based on the criteria detailed in the test specifications (see Appendix 3.5), which were developed on extensive analyses in the studies reported up to this point. The focus group (a Professor in Language Assessment and the researcher) chose several texts and subjected them to contextual parameters analysis to choose appropriate texts and discard others. Therefore, not only did the three texts chosen for test development reflect the textual features of first year university texts but they were equivalent to each other in terms of the textual features identified in the test specifications. For the rhetorical task, it was decided that an argumentative text would be more suitable for the purpose firstly because argumentative texts are assumed to contain more macropropositions and therefore may yield more main ideas questions (Urquhart and Weir, 1998), secondly because argumentative texts necessarily involve arguments and counter-arguments (contrasting ideas) with which arguments in other texts can be explicitly compared and contrasted. It was assumed that texts discussing advantages and disadvantages, pros and cons of an issue in a balanced manner would give sufficient arguments for and against an issue and be suitable for the tasks at all the three levels of careful reading. The texts were edited when necessary.

3.2.4.2 Task Design

Secondly, a variety of types of tasks were designed by the two members of the same group of experts, the researcher and the professor in Language Assessment, and analysed iteratively by a focus group of three experts in Language Assessment, to determine the most appropriate task types for our purposes. There were several considerations concerning the accurate and purposeful operationalisation of reading purposes and processes in the tasks which had to be addressed and these are discussed in relation to particular tasks below.

After the task types were decided on, the items were written with extra items to be eliminated as the tests were revised. The three versions of the test were iteratively revised and refined by the group of experts over several sessions. The experts worked in a focus group in which they discussed matters of relevance, representativeness, technical accuracy and practicality. The process stretched over several weeks and feedback from the experts was integrated in the process in an on-going fashion. Working on the basis of these refined test forms and extensive feedback, editing and revision, the focus group finalised three test tasks for each of the three reading processing levels. All the texts and tasks were reviewed once again for any oversights in typing and formatting.

The structure of the tests

As stated above, the test had three equivalent versions with three tasks in each version. In a counter-balanced manner, the same three texts were used for each version but each text was used for a different task in each version.

Table 3.4: Counter-balanced structure of the versions

	Text 1	**Text 2**	**Text 3**	**Versions**
Task S (sentence)	Task S/Text 1	Task S/Text 2	Task S/Text 3	*Version 1*
Task T (text)	Task T/Text 1	Task T/Text 2	Task T/Text 3	*Version 2*
Task I (intertextual)	Task I/Text 1	Task I/Text 2	Task I/Text 3	*Version 3*

The tasks

There is a brief description for each task below. It is not possible to reproduce the real test tasks here as they were for live tests that need to be kept confidential. Therefore, only some illustrative examples could be provided below. However, the examples are written solely for the purposes of exemplification to help the reader and they have not undergone any of the steps of the rigorous test construction process described above.

Task S (sentence level): Ability in local careful reading is a prerequisite for higher levels of reading; however, it never takes place in isolation in academic settings. For that reason, careful reading at sentence level is tested as embedded in a textual context; however it should be possible for the test taker to process the text locally, without having to form an integrated understanding of a main idea. Therefore, Task S questions were designed so that explicit information in one sentence could be linked with the question. Test takers were asked to choose an appropriate heading (subtitle) from a given list for each paragraph in the text. There was a clear semantic link between the headings and the first sentences (topic sentences) of each paragraph and the topic sentences of each paragraph were explicit enough to enable this matching. There was no lexical overlap between the topic sentences and the headings to prevent lexical match through scanning, and there were more headings than the paragraphs. Therefore, test takers were expected to read the first sentence of the paragraphs and to be able to choose a suitable heading. There were seven paragraphs to be matched with eight subtitles one being distractor to prevent guessing. The first question was given as an example thus there were six matches to be done. One point was assigned for each correct answer.

For example, the following paragraph in the main text titled '*Human Germline Engineering*' would be matched with the given heading, '*Is change safe?*'.

Human Germline Engineering

Is change safe?

It's not clear where human germline engineering will take us, but it's certainly clear that it could completely transform what it means to be human and alter the human species within a matter of centuries. Some say this is interfering in a realm that human beings should not be intruding into. Others say we don't have the wisdom to do this kind of manipulation or claim that it gives us too much power over future generations and could be physically damaging to them. There could be all sorts of accidents because we just don't know enough. This is what concerns people and this is why human germline engineering is such a difficult topic for us to deal with.

Task T (Text level): This task should necessitate the processing of the whole text forming a macrostructure of the main ideas. As Enright et al. (2000: 6) put it, such tasks 'require the reader to integrate and connect the detailed information provided by the author into a coherent whole'. As well as requiring the use of several components of knowledge base, tasks at text level might necessitate several cycles of reading and integrating information. The key issue here is that the test taker should not be able to complete the task by processing the text at ongoing mental representation level, which is where the reader integrates a few individual ideas, rather than conceptually integrating the information in the whole text and forming a macrostructure of the discourse.

The usual format to assess macrostructure formation is summary writing. However, not only is summarisation a task fraught with scoring problems

(Cohen, 1994a; Huhta and Randell, 1996), it also requires considerable writing performance. Therefore it is not the best possible format to assess 'the reading comprehension skill' per se. The document analysis in Section 3.2.1 also showed us summarisation tasks such as in TOEFLibt are problematic as well since they can be completed by processing the parts of the text, not necessarily by forming a macrostructure (see Cohen and Upton, 2006).

The final decision on the format was that the test takers would be given a list of statements summarising parts of the text (summary statements) but not necessarily on a paragraph basis; a statement might summarise one or more paragraphs. The summary sentences would be given in a mixed order and there would be two distractors to prevent guessing. Test takers were also asked to order the correct statements as the information in the text as accurate recognition of the organisation of information in a text is an important process in summarising a text. In doing this, test takers would presumably need to process all the information in the text, recognise the structure and organisation of it and eliminate the minor details and incorrect information. The distractor statements involved incorrect information or minor details that would not be included in a summary of the text. Therefore, the test taker would need to identify six informationally correct statements and put them in the order the information appeared in the text. The first one given as an example, there were five summary statements to be chosen. Each pair of correct summary statements in a correct order was assigned one point, totalling to five points maximum.

For example, for the paragraph above the following sentence could be given as the summary statement:

> Attempting to manage human evolution may be seen too ambitious for the reason that we cannot fully foresee the outcomes yet.

Task I (Intertextual level): This was the most challenging task to design since it involved an unprecedented attempt. The document analysis in Section 3.2.1 had shown us that although there were some tasks in certain widely-used language tests that required combination of information across texts, none of these tasks focused on reading comprehension processes only and the comprehension was assessed through writing as an end product. For example, in TOEFLibt, the task was writing an essay combining information from a lecture from the listening part and a text from the reading part. The essay is scored by taking into consideration several essay marking criteria, the content – comprehension output – being only a part of the score. Therefore, an appropriate, objectively-scoreable pen and paper task was needed to assess the ability to read across texts. The strongest suggestion from the literature was that in reading across texts, the first step is *to build semantic links between the texts and form a new mental representation* (see Section 3.2.3).The relations between the texts or parts of texts can be as varied as the relation between the parts of a single text (however under different textual organisations): elaboration, exemplification, comparison, contrast, cause and effect, problem and solution, arguing for (complementary information), arguing against (contradictory information), justification. The tasks to be performed on these relations could be identifying similar information, contrasting information, mapping information from one text onto another (by inserting paragraphs) for the purpose of providing support or counter-argument, identifying dissimilar point of views, organising information from two texts in a chronological order. For example, one of the tasks designed to operationalise combination of information during the initial stages of the study was as such:

'Imagine you have identified some further pieces of information that might usefully be combined with the information in the text 'Human Germline Engineering' to **support** or **oppose** some of the arguments raised in it. Where in the text can the following pieces of information be best used?' Match the mini-text letters with paragraph numbers.'

However, such tasks proved to be too ambiguous and difficult to design without leading to lack of clarity at the test development stage. It was finally decided that instead of focusing on the type of the relations between texts or parts of the texts, as an initial attempt and as the primary process of reading across texts, Task I should depend on solely identifying 'a relationship' – semantic relatedness – between the information from a part of a text and information from a part of another text. Therefore, Task I involved reading and understanding a main text and reading and understanding mini-texts extracted from other texts and matching the paragraphs in the main text with the mini-texts that involved related information to the paragraphs. There were seven paragraphs in the main text and six mini-texts. One being example, the test takers should match five mini-texts with the five related paragraphs among seven paragraphs in the main text. Each correct answer was assigned one point. To remind the reader, this process appeared as the initial step of documents model reading in the literature review and the present study. Although the task as it is cannot reach as far as testing whether a documents model is formed or not as source information is not included in the process, its strength is that it does not require any writing on the part of the test taker and can be quickly and objectively marked.

Example:

The paragraph above is related with the mini-text below in the sense that the mini-text supports the idea that human engineering might be harmful. The information in it could coherently be integrated with the main text paragraph:

Genetic causation is complex, with multiple genes interacting to create one outcome or behaviour, and single genes having multiple effects. When a long-term genetic effect may not show up for decades after the procedure was administered, parents will risk a multitude of unintended and irreversible consequences for their children.

However, a mini-text such as the one below, might not readily match with the text paragraph although both are on the same issue:

Enhancement interventions in human germline engineering are any interventions designed to produce improvements in human form or function that do not respond to medical needs. Human engineering of that sort is in much of the research agenda of contemporary biomedicine. The new procedures and technologies emerging from research laboratories and hospitals – whether mood-altering drugs, substances to boost muscle mass or selectively erase memory or gene therapyr– can as easily be used to 'enhance' the species as to ease or treat illness.

In this task, it was assumed that the reader would carefully read the main text and understand the issue discussed in it and carefully read the mini-texts to detect the relevance of it to a certain part in the main text. As is mentioned above, this was assumed to legitimately represent an initial step in intertextual reading; *building up of semantic links between the texts.*

3.2.5 Test takers' Verbal Protocol on Texts and Tasks

Verbal protocols are now widely used procedures to gather evidence on content relevance of test items to the ability to be measured. Cohen (1994b, 70) suggests that qualitative evidence on the congruence between 'the tester's presumptions about what is being tested and the actual processes that the test taker goes through' has to be established for any further interpretation that can be made on the test results. Such variables as text length and the nature of the task/questions are known to cause difficulty for test takers (Skehan, 1998). Ünaldi (2004) suggests that content related evidence on reading comprehension tests should be collected on text features, reading operations, text span covered and test taking strategies for a unified understanding of the cognitive and metacognitive processes that test takers use in responding to items.

In line with the suggestions above, as the last step of the test development process verbal protocol data were collected from six international students studying at the University of Bedfordshire, Department of Language and Communication; BA or post-graduate programs. On entry to the programs, these participants were required to present an IELTS score of 6.5 and therefore their level of English was advanced though there might naturally be variations among them. The researcher explained to the participants the purpose of the study and that they were expected to take the test and describe what they were doing in as much detail as possible. They were reminded to time themselves and read the instructions carefully. The participants took the test and they verbalised their thoughts as they responded to the items. However, in quite a few cases they made clear that they might need to concentrate more on the test tasks and they could not report simultaneously. In those cases, the researcher asked how the participants read and what parts of the text they read after they responded to the questions, i.e. the data were collected both by concurrent and retrospective think aloud procedure. The responses were grouped according to the types of careful reading; sentence, text and intertextual, and a ratio of comments on a particular process to all comments on the processes is given on task basis where possible. Other participant comments are presented as well. It was not possible to work on an item by item basis as the participants, especially in the second task, handled a few questions at a time and moved between the text and the questions frequently most of the time not wanting to make any comments. As the main aim in collecting these verbal protocols was to determine whether or not success in responding to the test items depended on predetermined reading purposes and processes as well as whether there were any technical problems in the tasks, observations by the researcher are also reported in relation to each task in the test.

3.2.6 Administration of the Test

After the tests had reached their final forms, they were administered to a large group of students for data collection on test performance. For the site of the test administration, an English medium Turkish university, Bogazici University, was chosen since the researcher had access to a large group of students in that institution. The researcher contacted the administration and the Testing Office of the School of Foreign Languages, Bogazici University for the necessary permissions and arrangements.[11] The administration agreed on the condition that the results from the test could be used for their own research purposes and the teachers could use the scores as a minor class assessment score. To all the teachers, the test specifications and a detailed explanation of the test were sent a week before the test administration. The researcher held meetings with the class teachers prior to the test administration and the test procedures were explained in detail. The test was administered together in three different

[11] The School of Foreign Languages of Bogazici University is a one year preparatory school where the students are given intensive skill-based EAP instruction before they start their studies at the departments. The instruction at the school emphasizes intensive academic reading and this is reflected in the EAP proficiency test of the school.

sittings to 817 students from all the four levels of prep school classes; beginners, pre-intermediate, intermediate and advanced. The students were between the ages 18 and 25. The class teachers administered the test and the researcher was present in all the three sittings for the teachers to resort to if need be.

The second group of data was collected from university students from year one to four (215 students), a relatively small sample group as the researcher had to make arrangements individually with the lecturers, most of whom found it difficult to spare time for the exercise.

The test was administered in three versions. As explained above, a 3X3 counter-balanced method was used matching each task to each text. Each version of the test contained all three texts but was associated with a different task in each version (sentence level, text, intertextual). Each of the three tasks in a version was therefore performed on a new text. This allowed us to try and control for text effect as far as possible, and to determine whether or not it was the different texts that affected difficulty of the tasks. The combination of tasks and texts is repeated here in Table 3.4:

Table 3.4: Counter-balanced structure of the versions

	Text 1	**Text 2**	**Text 3**	**Versions**
Task S (sentence)	Task S/Text 1	Task S/Text 2	Task S/Text 3	*Version 1*
Task T (text)	Task T/Text 1	Task T/Text 2	Task T/Text 3	*Version 2*
Task I (intertextual)	Task I/Text 1	Task I/Text 2	Task I/Text 3	*Version 3*

Each version was administered to approximately equal numbers of test takers in each sitting of test administration (i.e., Version 1: N=340, Version 2: N=347, Version 3: N=345). For example, if there were 30 students in an exam room, ten took Version1, ten took Version 2, and ten took Version 3. The test papers were colour coded according to the versions (Version 1 white, Version 2 blue, Version 3 yellow) to facilitate equal distribution of each version in each exam room. The distribution of the tests to the levels was as in Table 3.2:

Table 3.5: Distribution of the test takers to the levels

Level/Version (N: 1032)	Version 1	Version 2	Version 3
Prep beginner	41	40	30
Prep pre-intermediate	90	99	100
Prep intermediate	99	78	102
Prep advanced	41	53	37
University year 1	26	26	28
University year 2	10	14	9
University year 3	18	16	15
University year 4	15	21	16
Total	**340**	**347**	**345**

3.2.7 Statistical analyses of test results

The data from the test administration were subjected to classical test analysis procedures (central tendency measures, reliability and item analysis), analysis of variance (ANOVA) and Principle Component Analysis (PCA) and Principal Axis factoring (PAF). For each procedure, the details are explained below:

Measures of central tendency: For the measures of central tendency and dispersion, mean and standard deviation estimates were used

(Brown, 1996). These values and the Kolmogorov-Smirnof normality test were used to determine whether the data were normally distributed or not. In the Kolmogorov-Smirnof test, p values higher than 0.05 were taken as the indication of normally distributed data. The higher the p value, the closer the distribution to the normal distribution. In addition, skewness and kurtosis values, which indicate normality when they are equal to zero, were taken into consideration. The distribution is judged to be near normal if the skewness value is between –1e0 and +1e0. Kurtosis coefficients smaller than –1e0 are considered platykurtic (flat distribution) whereas coefficients larger than 2.0 are considered to be leptakurtic (overly peaked distribution). Normal Q-Q plots were also checked as further analyses (i.e. ANOVA) assume normality in the data. For the estimate of internal consistency, Cronbach's Alpha (α), which is based on the average inter-item correlation, was used.

Item analysis criteria: Item analysis was done to evaluate the effectiveness of individual test items. Traditionally two procedures, item facility - IF (or item difficulty) and item discrimination - ID, are employed (Brown and Hudson, 2002). Item-total correlation calculations (CITC) and reliability estimates for individual items (alpha if item deleted - AIID) were also analysed.

Item facility (IF) is determined as the proportion of correct responses to total number of items. Item facility is inversely related to the actual difficulty of any given item; the higher the difficulty, the lower the proportion of correct responses in the whole group of test takers (Henning, 1987). IF values range from 0 to 1.00 and Henning (1987) and Alderson et al. (1995) suggest that items which are as near to a facility value of 0.5 as possible should be selected to have a widespread scores in a test. In terms of rejecting the item as too difficult or too easy, the suggested rule of thumb is to reject items with IF less than 0.40 or more than 0.70 (Brown and Hudson, 2002). However, this decision must be closely related with the purpose of the test, and as the test under scrutiny here is a reading test geared at higher level academic reading ability, the limits for item rejection were set at 0.20 and 0.80 boundaries, following Green and Weir (1998).

Item discrimination index (ID) shows the degree to which an item discriminates between weak and strong examinees in the ability being tested (Henning, 1987). The groups of 'high-scorers' and 'low-scorers' are isolated as upper and lower third (sometimes as upper and lower 25%, 27%, or 33%). In order to calculate ID statistics, IF for the upper and lower groups are calculated separately (by dividing the number of examinees answering correctly in that group by the total number of examinees) and finally by subtracting the IF for the lower group from the IF for the upper group. Therefore, a discrimination index of 1.00 would be considered very good and 0.40 or above would be fairly high (Brown and Hudson, 2002). In the present study, item discrimination was analysed by dividing them into five groups according to their total test performance (by analysing item discrimination patterns - IDPs). For the analysis of item discrimination patterns, the groups are ranged from the lowest to the highest performing group and it is expected that the percentage of candidates answering a certain item correctly (IF) will increase from the lowest to the highest group systematically; weaker students responding to the item incorrectly and the good ones correctly. It is usually helpful to produce graphic representations of the way items perform across the five bands. If an item discriminates well between all the bands, we will see a line which moves from the bottom-left hand corner to the top-right-hand corner in the graph relatively similar to the one in Figure 3.1. On the other hand, it is clear in Figure 3.2 that the item is too difficult (the line does not reach the top-right-hand corner), and it does not discriminate well between the levels either, since firstly there is a dip at the point that corresponds to band 5 and secondly there is not much change in the slope of the line from band 3 to 6 except for the dip.

Figure 3.1: IDP graph for a favourable item

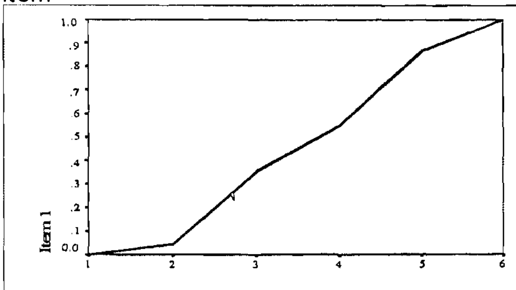
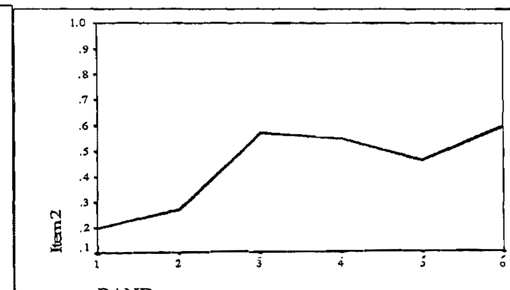


Figure 3.2: IDP graph for an unfavourable item



Item discriminability is also computed by looking at the correlation between an item and the total test/subtest score. This value is the correlation between the item and the score on the whole test/subtest minus the score on that item (corrected item-total correlation). A correlation of 0.20 and above is acceptable according to Green and Weir (1998). However, it should be born in mind that correlation is a function of sample size and ability range, and therefore may change with different samples (Henning, 1987).

Internal reliability estimates are additional data to evaluate the degree to which items fit together in a test, i.e., the test's homogeneity. Items that do not contribute to the test's overall reliability positively should be modified or rejected. This computation will tell us whether the test's internal reliability (alpha) would increase or decrease if the particular items were removed (alpha if item deleted). The statistical calculations were done using SPSS 17.

Comparison of the Group Means: In order to analyse the relationship between independent and dependent variables across groups, several comparison analyses were performed. When the data are continuous, independent, normally distributed and there are equal variances between the groups, Analysis of Variance (ANOVA) can be used. When these assumptions are not met, it is suggested that non-parametric alternatives should be used. The comparison of test taker scores across tasks and versions was done using *mixed between-within-subjects ANOVA*. The data from cognitive validity proforma were analysed through *Friedman Test*. In both cases, descriptive statistics were presented prior to mean comparisons.

Mixed Between-Within-Subjects ANOVA: Since there were two independent variables; task and version, a 3X3 factorial design was used in the analysis of the test scores. Therefore, it was possible to analyse whether the test takers performed differently across tasks, and across versions in which each task was coupled with a different text, and the interaction between task and version variables.

Friedman Test: For the analysis of the cognitive proforma (see Section 3.3), the non-parametric alternative of *one-way repeated measures ANOVA* was used since the data from the proforma were categorical and the same proforma was taken by all the test takers, which is there were no between subjects factor.

Post-hoc tests: For the test data, ANOVA's post-hoc test was used to specify where exactly the differences lay in the test score data. The non-parametric alternative, Wilcoxon Signed Rank Test was used for comparisons with the categorical cognitive proforma data. In order to reduce the risk of finding a difference by chance through repeated comparisons, Bonferroni adjustment to the alpha level was done dividing the alpha value (.05) by the number of comparisons, thus setting up a more stringent alpha level in each case.

Effect size: In order to analyse the degree to which the variables are associated with one another, the 'effect size' was calculated using the procedure 'partial eta squared'. Eta squared reveals 'the proportion of variance of the dependant variable that is explained by the independent variable' (Pallant, 2007). Values for eta squared ranges between 0 and 1 and are interpreted as below: .01= small effect, .06= moderate effect, .14= large effect.

Partial eta squared is an alternative technique known to overcome certain concerns with eta squared (See Pallant, 2007).

The Analysis of the Internal Structure of the Versions (PCA): Hatch and Lazaraton (1991e 490) state that Principle Component Analysis (PCA) is the technique used to 'determine whether it is possible to reduce a large number of variables to one or more values that will still let us reproduce the information found in the original variables. These new values are called *components* or *factors*. In the present study, to ensure accuracy in analysis, two methods, *Principal Component Analysis* (PCA) and *Principal Axis Factoring* (PAF) were computed using orthogonal (varimax), and oblique (direct oblimin and promax) rotation techniques for the investigation of the internal structure of the tests. While direct oblimin rotation technique did not reveal easily interpretable factor structures (e.g. some items not loading on any factors, some others loading on multiple factors), PCA with varimax and PAF with promax rotation yielded more interpretable and very similar results. For the sake of convenience, the results from PCA with varimax rotation are reported here and the ones from PAF with promax rotation are given in Appendix 4.6.

Principle Component Analysis was performed to analyse whether the internal structure of the test was consistent with the assumptions relating to three levels of careful reading. Together with ANOVA analysis, PCA might provide evidence for the congruence between the dimensions of the careful reading construct as reflected in the test specifications. The hypothesis was that the items putatively testing different operations (careful reading at sentence, text and intertextual levels) would load on different factors in PCA.

In this study, PCA was conducted following Hatcher's (1994, 1-56) suggestions. Hatcher describes PCA as a variable reduction procedure in which a set of observed variables are reduced into a smaller set of artificial variables called 'principal components' that will account for most of the variance in the observed variables. In language testing, it is seen as 'a way of discovering factors that underlie language performance and of testing the relationship among them' (Hatch and Lazaraton 1991, 489). The first component extracted in a PCA accounts for a maximal amount of total variance in the observed variables and the second component accounts for what is not accounted for by the first component and as such is uncorrelated with the first component. Resulting components will display varying degrees of correlation with the observed variables but will be uncorrelated with each other.

In a reading test for example, putatively different variables (e.g. skills) are expected to load on different components. If the variables load on the same component, we might assume that they function in a similar manner, and there is a strong possibility that there are no separate skills; they measure the same construct, undifferentiated reading ability. If, on the other hand, variables conceivably testing a certain skill load on a certain component while others load on a different component, we are led to think that reading ability is divisible as it is measured by that test (Green and Weir, 1998).

In conducting PCA, the first step is to perform an initial extraction of the components. The number of the components in the initial extraction is equal to the number of variables being analysed. However, for the subsequent analysis only a few of them will be retained for the interpretation. In this study, PCA was used to determine the number of the components/factors to be used in subsequent analysis. Principal Axis Factoring was then used to analyse the pattern of loadings on each factor. PAF as a method of Factor analysis, 'decomposes the score variances to isolate common variance and it also identifies unique variance (specific and error variance) accounted by each factor (Hatch and Lazaraton, 1991r. 492). Hatch and Lazaraton (1991r. 493) state that the main motivation to use factor analytic methods such as PAF following PCA is that 'factor analysis seeks the parsimony of the best common factors. It does not also try to explain specific and error variance (as does PCA).' Principal Axis Factoring was chosen to analyse the data since the assumption of normal distribution was violated and oblique (promax)[12] rotation was used to have uncorrelated factors. As both analyses yielded very similar results, only the results from PCA will be given in the next chapter. PAF analyses will be given in Appendix 4.6, as stated above.

To determine the number of components to be retained from PCA, the following criteria were taken into consideration:[13]

1. The eigenvalue-one criterion: An eigenvalue represents the amount of variance that is accounted for by a given component. This criterion suggests that any component with an eigenvalue greater than 1.00 should be retained. Since each variable contributes one unit of variance in the data set, any component that displays an eigenvalue that is more than 1.00 accounts for a greater amount of variance than

[12] Rotation is a linear transformation that is performed on the factor solution to make the interpretation easier. Varimax rotation produces uncorrelated components and maximises the variance of a column of the factor pattern matrix (Hatcher 1994, 28).
[13] In the present study, 'communalities' (percent of variance in a variable that is accounted for by the retained components) are also considered. Variables are expected to have a communality value of .3 or more.

itself. Components with eigenvalues less than one are viewed as trivial and are not retained.

2. The scree-test: A scree-plot displays eigenvalues against components. When there is a large break in the curve and it starts to flatten out, the components after the break are assumed to be unimportant and are not retained.
3. Proportion of variance accounted for: It is suggested that we may retain components that account for at least 5% of the total variance. Alternatively, researchers might retain enough components so that a cumulative percent of 70% is attained.[14]
4. Interpretability criteria: The basic question here is whether the retained components have substantive meaning and whether our interpretation of the components makes sense in terms of what is known about the constructs under investigation (Hatcher 1994, 22-26).
5. Parallel Analysis: This is a newly recognised technique that calculates the average eigenvalues for a specified number of randomly generated samples (Pallant, 2007). The generated eigenvalues are compared to the actual eigenvalues in the analysis and if the actual values are greater than the generated values, the factor is retained and if not, it is rejected. Parallel analysis in this study was performed using web-based software Monte Carlo PCA for Parallel Analysis developed by Watkins (2000, in Pallant 2007).

[14] However, Green and Weir (1998) warn us that when our data are of individual items on a 0/1 scale, the scale for any correlations is very restricted. Therefore, in our case too, we may expect to find lower cumulative percent of variance accounted for by the extracted components.

In the present study, the data were analysed taking primarily the first criterion, eigenvalue-one rule, into consideration. Then, Parallel Analysis was performed. After the number of components to be retained was decided, which generally corresponded to the number of components with eigenvalues over 1.00, the data were subjected to PCA with 'varimax rotation'. The variables that have high loading on one particular component were determined. The loading is usually considered sufficient when it is at least .300. Variables are expected not to have high loadings on more than one component. The resulting component structure was ideally expected to be with three components that account for three types of careful reading; sentence (propositional level), text (situational model) and intertextual (documents model) levels. To determine whether the data were adequate for the analyses, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was used. This measure compares the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients. A KMO below 0.50 is usually considered unacceptable. Another indicator of the strength of the relationship among variables, namely Bartlett's test of sphericity, was also used to check whether the observed significance level is small enough (SPSS Base 10.0 User's Guide,1999). SPSS version 17 was used in the analysis of the data.

3.3 Research Question 3) Is cognitive processing different in the various types of careful reading?

Using the information from the earlier small scale diary study, earlier review of the cognitive processing literature and the questionnaire surveys (Green et al. 2008 and Weir et al. 2009), a tool named 'cognitive validity proforma' was designed to elicit reading operations that a test taker might use for each task (Task S, Task T, Task I) and the amount of text he or she had to read to answer the items. This proforma was piloted and trialled in an earlier study of reading behaviour of 352 test takers taking the IELTS test (see Weir et al. 2009). In the current study of cognitive processing in careful reading tasks, the proforma was reformatted to enable candidates to feed back on each section of the test they were taking (see Appendix 3.6). Through cognitive validity proforma, data were collected on the reading operations the test takers used when answering the test questions (Q1-Q8) and they also reported the amount of text they processed (Q9-Q14) giving us a profile of cognitive processes they used in the completion of the test. All the students taking the tests were asked to complete the proforma immediately after each task was attempted. A total of 931 - over 300 valid responces to each version - were obtained (i.e., Version 1: N=319, Version 2: N=301, Version 3: N=311). The data from the cognitive validity proforma were analysed using descriptive statistics, ANOVA (Friedman Test) and planned comparisons (Wilcoxon Signed Rank Test), the procedures which are explained above.

This chapter has presented the methodological details of the studies designed to probe the three research questions formulated for the investigation of careful reading process in academic settings. The next chapter will present the results of the studies and related discussions in the order of research questions.

CHAPTER 4

RESULTS and DISCUSSION

As discussed in the previous chapters, the aim of this study is to establish the types of careful reading in academic settings by investigating manifestations of different types of careful reading as affirmed by students through questionnaires, reading diary studies, interviews and also as they surfaced in the test and associated cognitive validity pro forma. This chapter presents the results from those studies as they relate to the three research questions. Discussion on the issues raised will be presented immediately after the results.

4.1 Research Question 1) What are the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution?

The first research question was investigated through a large scale questionnaire, reading diary study and follow-up interviews.

4.1.1 The Questionnaire Study and Statistical Analysis

Below are the statistical analyses of the main administration of the questionnaire. The results will be reported as they apply to the research questions of this study. Questionnaire responses are reported grouping the participants into four categories: English as a first language (EL1), English as an additional language (EAL) and Year 1 and Year 2 students.

4.1.1.1 Demographic Features

Table 4.1 shows the distribution of the participants according to gender, age and regional backgrounds and Table 4.2, academic stage and Table 4.3, subject areas.

Table 4.1: Gender, age and regional distribution of the questionnaire participants (Weir et al, 2009)

Male/ Female	N	%	Age Range	N	%	Region	N	%
M	227	29.6	18-22	427	55.7	UK	287	37.6
			23-29	178	23.2	EU	135	17.7
F	537	70.1	30-39	92	12.0	Other	342	44.8
			40+	69	9.0			
N	764			766			764	

Table 4.2: Level and stage of the questionnaire participants (Weir et al, 2009)

Level	N	%	Year	N	%
Undergraduate	642	84.4	1	518	67.7
Postgraduate	118	15.6	2	230	30.3
			3	15	2.0
N	761			758	

Table 4.3: Subject areas (Weir et al, 2009)

Subjects	EAL		EL1		Subjects	EAL		EL1	
	N	%	N	%		N	%	N	%
Advertising, Management, PR	94	20.4	12	4.1	Language and Communication.,TEFL	71	15.4	7	2.4
Art & Design	2	0.4	1	0.3	Law	10	2.2	7	2.4
Biology, Biomedical Science	9	2	4	1.4	Leisure, Tourism, Sports, Management	11	2.4	3	1.0
Business & Finance	137	29.8	18	6.1	Media Arts	24	5.2	10	3.4
Computer and Information Science	16	3.5	15	5.1	Psychology	16	3.5	28	9.5
Education Studies	5	1.1	66	22.4	Social Sciences, Social Work	7	1.5	42	14.2
Healthcare, Nursing	7	1.5	40	11.2	Sport & Exercise	3	0.7	31	10.5
Human Resource Management	36	7.8	16	5.4					

EAL: English as an additional language EL1: English as a first language

There are significantly more female participants in the sample than male and the largest age range group is 18-22 as expected in a study focusing on first year students. The number of participants from other age groups is satisfactory. 37.6% of the participants are of British origin, 17.7% are from European countries and 44.8% are from non-European origin, resulting in a good distribution among British and EU and international students. There is a predominance of undergraduate students (84.4%) and among undergraduate and postgraduate students, most are in their first year (67.7%). In the EAL population, a high proportion (66.2%) is between 18 and 22 years old. In fact, a fairly similar proportion of the EAL and the EL1 sub-groups, 81r.9% and 88.2% respectively, are studying here at undergraduate level. Analysis of the year of study category across our EAL and EL1 groups shows high proportions of students (70.1% and 64% respectively) in their first year of study at the University. There is a group of second year students (30.3%) as well, large enough to make comparisons with respect to the changes that may occur in aspects of their academic reading.

A broad range of subject areas is represented in the sample. Business and Finance (29.8%), Advertising Marketing and Public Relations (20.4%) and Language and Communication (15.4%) are the main subject areas for EAL students. EL1 students marked Education Studies (22.4%), Social Sciences and Social Work (14.2%) and Sport and Exercise Science (10.5%) as their area of study.

In the sample, 43 languages were represented; English (38.9%) and Chinese (38.4%) being the most numerous, with European languages other than English represented by 14.2% of the sample. The first and second year student groups are well matched proportionately across first languages, dominated by EL1 (Year 1 and Year 2 at 36.5% and 44.8% respectively) and Chinese L1 (36.5% and 33.5%). Of the EAL students, 66.5% are from outside the UK, 27.9% from Europe, and 5.6% from the UK. Among the 298 students in the EL1 group, 87.6% are UK nationals, five European and 10.7% from outside the UK and Europe.

The number of students in the categories shows a useful distribution for the purposes of this study. The preponderance of female students is also expected since The Higher Education Statistic Agency (HESA) statistics notes that 58% of the students in UK universities in the academic year 2003/2004 are females. Therefore, we have a large group of first year and second year students in undergraduate and postgraduate studies pursuing a range of subjects.

4.1.1.2 Reading Purposes, Skills and Problems

The key areas of interest in the reading experiences of the students in the first year of their courses at a British university are the purposes of reading – how they read in their courses; cognitive processing – what they do when they read for assignments, and the difficulties they experience. The analysis of the questionnaire is done based on the 5 point Likert scale items: *5 definitely agree, 4 mostly agree, 3 neither agree nor disagree, 2 mostly disagree, 1 definitely disagree*. The response tendencies on each item are reported in modes; the most common selection made by the participants. In order to indicate the strength of the responses, the categories *definitely agree* (D) and *mostly agree* (M) are collapsed to D&M in the statistics.

Table 4.4 shows how EAL, EL1 and Year1, Year 2 students ranked four (15-1i8) *purposes of reading* in the order of importance. Among the four reading purposes, 15 refers to 'search reading'. The rest are careful reading purposes: 16 'reading at local level', 17 'text level' reading and 18 refers to reading 'multiple-texts'. The numbers in parentheses indicate the rank orders (r/o) of the reading agreement strengths. It is clear from the table that all the reading purposes are marked as important in general. In particular, for both EAL and EL1 students '*searching texts to find information*' and '*understanding meaning of a text as a whole'* are the first two important purposes. When Year 1 and Year 2 students are compared, the students ranked search reading still as the most important skill. However, Year 2 students ranked *reading across texts* as the second most important purpose.

Table 4.4: Purposes for reading (Weir et al, 2009)

The following purposes for reading are important on my course:	**EAL** D&M (r/o)	**EL1** D&M (r/o)	**Year 1** D&M (r/o)	**Year 2** D&M (r/o)
15.Searching texts to find information for assignments and exams	**87.7% (1)**	**95.6% (2)**	**91.2% (1)**	**90.7% (1)**
16. Basic comprehension of main ideas	79.6% (3)	90.1% (4)	**83.5% (2)**	84.5% (3)
17. Understand meaning of text as a whole; how main ideas and details relate to each other and author's purpose	**80.7% (2)**	**97.5% (1)**	82.2% (4)	82.4% (4)
18. Integrating information from different texts for use in assignments, exams	78.6% (4)	91.1% (3)	83% (3)	**85.5% (2)**

EAL: English as an additional language
EL1: English as the first language
Year 1: the first year students;
Year 2: the second year students
D&M: definitely agree and mostly agree
r/o: rank order

Table 4.5 reports the reading strategies the respondents reported using while preparing assignments. In the whole list of strategies, there are 16 items covering pre-reading strategies, cognitive operations taking place during reading including both expeditious and careful reading operations as well as note-taking (32) and comprehension monitoring strategies (30). The table reports the strategies pertaining to careful reading only (26-34) since the focus of this study is specifically on careful reading. It summarises the responses in terms of strength of agreement (definitely and most strongly agreed) with each item across the EAL and EL1 groups and Year 1 and Year 2 groups with average rank orders given in parentheses (1-16; 1 being the most important, 16, the least).

Table 4.5: Responses on ways of reading for assignments across sub-groups (Weir et al, 2009)

***How do I read for assignments* (16 reading strategies)**	**EAL D&M (r/o)**	**EL1 D&M (R/O)**	**Year 1 D&M (r/o)**	**Year 2 D&M (r/o)**
26. If I do not know the meaning of a word in a text, I try to work out its meaning	63.8% (10)	**77.5% (7)**	**69.9% (8)**	66.6% (9)
27. I read a text slowly all the way through even if some parts do not seem relevant to my assignment	33.4% (16)	30.7% (16)	33.4% (16)	32.6% (16)
28. I read slowly only those sections of a text I have marked as relevant when going through it quickly before	**67.6% (6)**	63.7% (11)	66.7% (11)	65.7% (11)
29. While reading I try to relate content to what I know already and judge its value	**67.4% (7)**	**78.9% (5)**	**71.2% (5)**	**75.0% (3)**
30. I look back at previous parts of the text to check meaning	64.4% (9)	75.6% (8)	69.8% (9)	66.3% (10)
31. I try to understand how the text is organized: how the ideas and details connect with each other	61.9% (12)	62% (12)	66.6% (12)	52.0% (15)
32. I make notes on relevant points from the text as I go along	66.4% (8)	**78.7% (6)**	**70.3% (7)**	**72.3% (5)**
33. I integrate information from the text I am reading with information from other texts I have already read	62.4% (11)	**79.1% (4)**	69.3% (10)	**69.7% (6)**
34. I read critically to establish and evaluate the author's position on a particular topic	51.4% (15)	59% (13)	54.3% (15)	55.3% (13)

In general, EAL students did not rank careful reading strategies as most important. Among 16 strategies, they marked 28 and 29 as relatively important (6th and 7th respectively); the former indicating local careful reading following selective reading and the latter, *use of content knowledge* for the evaluation of the new information. EL1 students on the other hand, ranked *integrating information across texts* as the fourth important skill, and the use of *content knowledge* as the fifth important skill. While for Year 1 students *using content knowledge* (5th) and *taking* *notes* (7th) seem important, Year 2 students also ranked *integrating information across texts* (6th) as relatively important. There were slight differences between EAL and EL1 students, the latter giving slightly more importance to vocabulary (26), background knowledge (29), taking notes (32) and reading critically (34); the former, marking texts and rereading the important parts (28). Year 2 students ranked background knowledge (29), taking notes (32) and reading critically (34) as slightly more important, and Year1 students ranked understanding the organisation of information (31) as more important.

One striking difference between the groups is on the importance of integrating information across texts (33). Considering the order of importance the respondents ranked 16 reading strategies, EL1 and Year 2 students ranked it as more important than EAL and Year1 students (4th, and 6th versus 11th and 10th, respectively).

4.1.1.3 Discussion on the Questionnaire Study

The questionnaire study helped us establish that among other reading purposes and strategies, careful reading at several levels is perceived as an important skill for academic study. It is also note-worthy that native speakers of English and more experienced students perceived reading multiple texts to integrate information as relatively important in comparison to non-native speakers and first year students. It may be inferred that the importance of the strategyh or the skill as it were – may change according to the language background or the educational experience of the students. It seems that *integrating information across texts* gains importance as the academic experience of the students' increases (see Carlson 2001). Once it has been established that in academic settings students need to process texts in a variety of ways including local comprehension, comprehending a text in its entirety and reading more than one text to integrate information, we can look in more detail at how processing at different levels takes place and what cognitive operations are involved in each level. The diary study below will enable us to study the reading activities of students reading for assignment preparation over a period of time and to see the relations between reading purposes and cognitive operations.

4.1.2 The Reading Diary Study

The reading diary study will be reported in the order of the sections in the diary forms. Where we need to analyse *reading purposes* and *cognitive processes*, we will refer to not only careful reading strategies but to all and we will analyse the place of different types of careful reading in the entire process of reading for assignment preparation. In analysing the Likert scale responses, the percentages of positive responses, i.e. *definitely agree* and *mostly agree* are summed and compared with neutral (*neither agree nor disagree*) and negative responses (*mostly disagree* and *definitely disagree)* in the data. The purposes and processes endorsed by 50% or more will be focused on.

4.1.2.1 Contextual Parameters

The first section of the reading diary form was on contextual parameters including the source of the text, the amount of time spent on it, mode and location. The participants mostly read books, book chapters or journal articles (38.7% and 34% respectively). 11.3% of the material was sourced from an internet site, 8% was reports and 8% was other kinds of material such as theses or unpublished manuscripts. The readers mostly read the texts only once (75.8%). Most reading was carried out using paper-based material rather than on screen (14.5%). Reading was most often done at home (51i.6%) or at the office (46.8%). The location was a library in just 1.6% of the cases.

4.1.2.2 Reading Purposes and Cognitive Processes

Table 4.6 shows the frequency of positive responses for Section 2: *reading purposes*. The pattern of positive responses suggests the importance to the participants of integrating information across texts in building an understanding of a topic. These included *13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.* (66.1%). *11. Evaluate the writer's* *ideas on the topic.* (54.8%), *3. Find support or opposition to an idea I had in mind.* (54.8%) and *12. Compare the writer's point of view with that of the other writers* (50.0%). The following items were also endorsed by a majority of participants; *10. Form an understanding of how the main ideas in the text relate to each other* (56.5%) and *4. Find specific information to answer a question* (51s6%). 10 typically describes the need to comprehend a whole text and the ideational relations within. 4, on the other hand, denotes a search reading purpose to extract relevant information.

Table 4.6: Section 1 Reading purposes by percentage (% endorsing)

Reading Purposes	**percent**
13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.	66.1%
10. Form an understanding of how the main ideas in the text relate to each other.	56.5%
11. Evaluate the writer's ideas on the topic.	54.8%
3. Find support or opposition to an idea I had in mind.	54.8%
4. Find specific information to answer a question.	51.6%
12. Compare the writer's point of view with that of the other writers.	50%
9. Understand and remember as much information as possible, but only from relevant parts of the text.	43.5%
2. Locate parts of the text that would be worth reading later.	41.9%
8. Understand and remember as much information as possible from the text as a whole.	40.3%
5. Get basic understanding of just the main idea(s) in the text.	37.1%
1. Find specific words, phrases or figures.	30.6%
7. Check my comprehension of a specific part of the text.	22.6%
6. Decide whether the text would be worth reading.	16.1%

Table 4.7 below shows the results for Section 3: cognitive processes. In the majority of cases, the readers endorsed *14. I used my knowledge of the topic to help me to understand the text* (87.1%). The popularity of 9. *As I read, I tried to form a summary of the ideas* in my mind (71%); *7. As I read, I tried to connect information from different sentences in the text* (67.7%) and *10.I tried to combine information from this text with*

information from other texts (61.3%) indicate an important role for careful reading processes in assignment preparation. The re-reading, processing the text locally where important information is situated, was also endorsed by 53.2% (*12. I read the relevant parts of the text again*).

Endorsed by 87.1%, 14 clearly shows the importance of previously acquired topic knowledge. 9 indicates that processing a whole text in order to reduce it to its macrostructure – careful reading at text level – is an important process for these readers. The readers also read carefully at local level (7) processing meaning in individual sentences and combining information incrementally as they read, and they re-read carefully at local level again. However, participants also frequently reported engaging in search reading processes such as *1. I used the headings, titles, contents page or index to locate information quickly* (62.9%); *5. I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs* (56.5%), and *13. I used my knowledge of how texts like this are organised to find parts to focus on* (51.6%). It is notable that there are fewer endorsements of 8 (*I read the text slowly all the way through from beginning to end*, 41o9%) than 5 (*I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs*, 56.5%).

Table 4.7: Section 2 Cognitive Processes (% endorsing)

Cognitive Processes	percent
14. I used my knowledge of the topic to help me to understand the text.	87.1%
9. As I read, I tried to form a summary of the ideas in my mind.	71%
7. As I read, I tried to connect information from different sentences in the text.	67.7%
1. I used the headings, titles, contents page or index to locate information quickly.	62.9%
10. I tried to combine information from this text with information from other texts.	61.3%
5. I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs.	56.5%
12. I read the relevant parts of the text again.	53.2%
13. I used my knowledge of how texts like this are organised to find parts to focus on.	51.6%
8. I read the text slowly all the way through from beginning to end.	41.9%
15. I used my knowledge of the world to help me to understand the text.	38.7%
4. I read the key parts of the text such as the abstract, introduction and conclusion.	32.3%
6. I looked for the parts of the text the writer indicates to be important.	30.6%
2. I looked quickly for the words relating to the topic of my assignment.	30.6%
3. I quickly matched words that appeared in the question with the same words in the text.	17.7%
11. I looked at other texts while I was reading this one.	16.1%

4.1.2.3 Fast Reading versus Slow Reading

A division appeared to occur in our data based on the number of pages of text that participants reported reading in a session. It was clear in the reading diary forms that sometimes the readers processed a long text in relatively short time and vice versa. On this basis, we make a distinction between relatively fast and slow reading sessions with the former involving reading a page of text (assumed on the basis of the five texts we explored to be between 350 and 450 words) in one or two minutes. In *slow reading* it typically took the students around six minutes to read each page (i.e. around 60 to 80 words per minute). All the participants provided a balance of fast and slow reading data although one participant indicated on her diary forms that she read fast in almost all (92%) of the sessions she reported. The distribution of reading speeds is set out in Table 4.8. This shows that 27 of the 62 reading sessions involved slow reading, compared with 35 spent in fast reading.

Table 4.8: Reading speed by type (fast or slow reading)

Reading Type	Fast	Slow
Number of reading sessions in data	37	25
Minutes per page		
Mean	1.18	6.04
Std. Dev.	0.56	3.18
Minimum	0.15	3.00
Maximum	2.40	15.00

It is evident from Table 4.9 that the major purpose for fast reading is *2. Locate parts of the text that would be worth reading later* (59.8%). This item was only endorsed on 11.1% of the *slow* reading forms. In each of the four most popular reading purposes for *fast* readers (*2. Locate parts of the text that would be worth reading later*; *3. Find support or opposition to an idea I had in mind*; *4. Find specific information to answer a question*; *13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere*), the sessions are dominated more by the purposes of finding and locating information previously conceived in mind; the reader has a specific topic in mind either from earlier reading or from the given assignment and he or she is targeted to find information on the topic.[15]

[15] It is notable that 6 (*Decide whether the text would be worth reading)* is not a popular selection in the *fast* reading data.

Table 4.9: Fast reading: Reading purposes

Fast Reading - Reading Purposes	percent
2. Locate parts of the text that would be worth reading later.	59.5%
3. Find support or opposition to an idea I had in mind.	56.8%
4. Find specific information to answer a question.	56.8%
13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.	51.4%
9. Understand and remember as much information as possible, but only from relevant parts of the text.	48.6%
5. Get basic understanding of just the main idea(s) in the text.	43.2%
1. Find specific words, phrases or figures.	37.8%
10. Form an understanding of how the main ideas in the text relate to each other.	32.4%
11. Evaluate the writer's ideas on the topic.	29.7%
12. Compare the writer's point of view with that of the other writers.	27%
6. Decide whether the text would be worth reading.	27%
7. Check my comprehension of a specific part of the text.	16.2%
8. Understand and remember as much information as possible from the text as a whole	2.7%

Among the cognitive processes (see Table 4.10) involved in *fast* reading, the most commonly endorsed was *14. I used my knowledge of the topic to help me to understand the text.* (91.9%). Background knowledge greatly facilitates efficient text processing and appears to have played a key role in allowing the readers to read a text quickly and selectively. Efficient use of textual features (*1. I used the headings, titles, contents page or index to locate information quickly*: 86.5% and *13. I used my knowledge of how texts like this are organised to find parts to focus on*: 67.6%) also seems a feature of fast reading in this context. However, these readers also endorsed items such as *5. I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs* (57.6%) and *12. I read the relevant parts of the text again,* which suggests that our participants did read carefully where they identified useful information.

Table 4.10: Fast reading: Cognitive Processes

Fast Reading - Cognitive Processes	**percent**
14. I used my knowledge of the topic to help me to understand the text.	91.9%
1. I used the headings, titles, contents page or index to locate information quickly.	86.5%
5. I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs.	67.6%
13. I used my knowledge of how texts like this are organised to find parts to focus on.	67.6%
9. As I read, I tried to form a summary of the ideas in my mind.	59.5%
12. I read the relevant parts of the text again.	54.1%
10. I tried to combine information from this text with information from other texts.	51.4%
7. As I read, I tried to connect information from different sentences in the text.	51.4%
2. I looked quickly for the words relating to the topic of my assignment.	45.9%
15. I used my knowledge of the world to help me to understand the text.	40.5%
6. I looked for the parts of the text the writer indicates to be important.	35.1%
4. I read the key parts of the text such as the abstract, introduction and conclusion.	32.4%
3. I quickly matched words that appeared in the question with the same words in the text.	27%
8. I read the text slowly all the way through from beginning to end.	27%
11. I looked at other texts while I was reading this one.	18.9%

The three most frequently endorsed purposes for *slow* reading (see Table 4.11) are *8. Understand and remember as much information as possible from the text as a whole; 10. Form an understanding of how the main ideas in the text relate to each other; 13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.* (all 77.8%), clearly indicating the need to comprehend the whole text in detail and relating the content to other texts. The popularity of *11. Evaluate the writer's ideas on the topic* (74.1%), *12. Compare the writer's point of view with that of the other writers* (70.4%) and *3. Find support or opposition to an idea I had in mind* (59.3%) also underlines the importance of forming intertextual, ideational relationships across texts in the *slow* reading approach.

Table 4.11: Slow reading: Reading purposes

Slow Reading - Reading Purposes	**percent**
8. Understand and remember as much information as possible from the text as a whole.	77.8%
10. Form an understanding of how the main ideas in the text relate to each other.	77.8%
13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.	77.8%
11. Evaluate the writer's ideas on the topic.	74.1%
12. Compare the writer's point of view with that of the other writers.	70.4%
3. Find support or opposition to an idea I had in mind.	59.3%
4. Find specific information to answer a question.	44.4%
7. Check my comprehension of a specific part of the text.	29.6%
9. Understand and remember as much information as possible, but only from relevant parts of the text.	25.9%
1. Find specific words, phrases or figures.	22.2%
5. Get basic understanding of just the main idea(s) in the text.	18.5%
2. Locate parts of the text that would be worth reading later.	11.1%
6. Decide whether the text would be worth reading.	0%

The cognitive processes that were most characteristic of *slow* reading (see Table 4.12) included the following: *7. As I read, I tried to connect information from different sentences in the text* (92.6%); *9. As I read, I tried to form a summary of the ideas in my mind.* (85.2%); *14. I used my knowledge of the topic to help me to understand the text* (81.5%) and *10. I tried to combine information from this text with information from other texts* (74.1%). The cognitive processes endorsed most when the readers read using relatively more time noticeably relate to careful reading processes when a reader processes a text sentence by sentence (propositional level) and when they incrementally combine the ongoing information to form a macrostructure (text level) and also when a reader integrates information across texts (multiple text level).

Table 4.12: Slow reading: Cognitive processes

Slow Reading - Cognitive Processes	Percent
7. As I read, I tried to connect information from different sentences in the text.	92.6%
9. As I read, I tried to form a summary of the ideas in my mind.	85.2%
14. I used my knowledge of the topic to help me to understand the text.	81.5%
10. I tried to combine information from this text with information from other texts.	74.1%
8. I read the text slowly all the way through from beginning to end.	63%
12. I read the relevant parts of the text again.	51t9%
5. I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs.	40.7%
13. I used my knowledge of the world to help me to understand the text.	37%
4. I read the key parts of the text such as the abstract, introduction and conclusion.	33.3%
1. I used the headings, titles, contents page or index to locate information quickly.	29.6%
13. I used my knowledge of how texts like this are organised to find parts to focus on.	29.6%
6. I looked for the parts of the text the writer indicates to be important.	25.9%
11. I looked at other texts while I was reading this one.	11.1%
2. I looked quickly for the words relating to the topic of my assignment.	7.4%
3. I quickly matched words that appeared in the question with the same words in the text.	3.7%

4.1.2.4 Correlations between Reading Purposes and Cognitive Processes

Having analysed reading purposes and cognitive processes (Section 4.1.2.3) separately, we next sought evidence for the relationship between the two sections of the diary form. This would provide confirmatory evidence on the link between reading purposes and cognitive processes as reported by the participants in the study. Table 4.13 below shows the significant correlations ($p<.01$) between reading purposes and cognitive processes employed during different types of reading. For ease of reference, the entries are referred to in abbreviated form.

In general, the activities geared to expeditious reading are linked with the processes that define skimming, search reading and scanning. The first seven reading purposes correlate significantly with first six cognitive activities as well as knowledge on text organisation (13) and knowledge of the world in general (15). Looking at other texts (11) and re-reading the relevant parts (12) also appeared somewhat correlated with the first seven reading purposes. When the reader wants to focus on certain parts of the text (9) several processes from matching of words and looking for relevant information (scanning and search reading) to using textual features, using background knowledge and reading carefully seem to be involved.

Reading Purposes	Cognitive Processes	(Spearman's rho)
1. Specific words:	3. Matched question words	.686
	11. Looked at other texts	.582
	15. Used knowledge of the world	.574
	2. Words relating to topic	.530
	6. Parts writer indicates	.528
	12. Read relevant parts again	.502
	1. Titles etc to locate quickly	.417
	5. Slowly carefully relevant parts	.400
	4. Key parts	.379
	13. Used knowledge of organisation	.367
2. Locate parts to read later:	1. Titles etc to locate quickly	.522
	15. Used knowledge of the world	.350
	13. Used knowledge of organisation	.310
	2. Words relating to topic	.309
	3. Matched question words	.286
	11. Looked at other texts	.253
3. Find support or opposition:	5. Slowly carefully relevant parts	.345
4. Find specific info:	5. Slowly carefully relevant parts	.334
	3. Matched question words	.289
5. Get basic understanding:	15. Used knowledge of world	.499
	1. Titles etc to locate quickly	.437
	6. Parts writer indicates	.363
	13. Used knowledge of organisation	.306
6. Decide whether worth reading:	1. Titles etc to locate quickly	.392
	2. Words relating to topic	.303
7. Check comprehension:	12. Read relevant parts again	.538
	11. Looked at other texts	.357
	3. Matched question words	.284
8. Understand and remember as much as possible:	8. Slowly beginning to end	.491
	9. Form summary in mind	.491
	7. Connect information across sentences	.253
9. Understand and remember bits:	5. Slowly carefully relevant parts	.599
	12. Read relevant parts again	.453
	15. Used knowledge of world	.447
	3. Matched question words	.363
	11. Looked at other texts	.341
	7. Connect information across sentences	.336
	6. Parts writer indicates	.316
	13. Used knowledge of organisation	.301
	4. Key parts	.284
	2. Words relating to topic	.257
10. Understanding of main idea relationships:	9. Form summary in mind	.436
	10. Combine information across texts	.393
	7. Connect information across sentences	.287
	8. Slowly beginning to end	.260
	12. Read relevant parts again	.254
11. Evaluate writer's ideas:	10. Combine information across texts	.502
	7. Connect information across sentences	.341
	9. Form summary in mind	.322
	8. Slowly beginning to end	.287
	6. Parts writer indicates	.260
12. Compare writers' points of view:	10. Combine information across texts	.586
	9. Form summary in mind	.393
	7. Connect information across sentences	.331
	8. Slowly beginning to end	.301
	14. Topic knowledge	.294
	12. Read relevant parts again	.284
13. Judge this text in relation to other texts:	10. Combine info across texts	.793
	7. Connect information across sentences	.484
	14. Topic knowledge	.460
	9. Form summary in mind	.325
	12. Read relevant parts again	.252

When careful reading purposes were analysed, reading and understanding a text (*8. understanding and remembering as much information as possible from a text as a whole*) seem to be related with reading sentence by sentence (*7 connecting information across sentences*) to the end (*8 reading slowly beginning to end*) forming a macrostructure of the text (*9 forming summary in mind*r). A similar purpose, 10, (*10 forming an understanding of how the main ideas in the text relate to each other*), correlates with the same processes as well as re-reading (*12 reading relevant parts again*) and importing information from other texts (*10 combining info across texts*). When the readers read criticically, i.e. when they wanted to evaluate (*11 looking at other texts*) and compare the writer's ideas (12), and when they wanted to judge how the text they were reading was related to other texts they had read in terms of supporting or developing ideas they had found elsewhere (13), it becomes important to combine information from the text at hand with information from other texts. These reading purposes also involve processing a whole text (8 and 9) sentence by sentence (7) but they require re-reading (12) and resorting to topic knowledge (14), too.

For all the purposes listed above, the readers seem to have read sentence by sentence, building the macro-structure of the text incrementally. With the exception of *8*, all the other purposes are related with the cognitive process *10* (*trying to combine information from this text with information from other texts*), which shows that previously attained information is important in more critical/evaluative reading attempts. Rereading (*12; reading the relevant parts of the text again*) is resorted to while *forming an understanding of how the main ideas in the text relate to each other* (*10*), *comparing the writer's point of view with that of the other writers* (*12*) and *judging how this text was related to other texts that were previously read* (*13*). However, topic knowledge (14) is only significantly (p<.01) correlated to *12* and *13*.

4.1.2.5 Storing Information, Difficulties and Usefulness

This part of the diary study gave us some idea on the ways the readers stored information for further use in assignment preparation, the difficulties they had and their perceptions on the usefulness of the texts they chose. Storing information, the readers either made notes on relevant points from the text (mean = 2.95) or highlighted the text as they read (mean = 2.84) (see Table 4.14). It is also interesting that they often reported remembering where the relevant information was (mean = 2.71). In relatively fewer cases, the participants reported copying extracts to a separate document (mean = 1.82) or typing directly into the body of their assignment (mean = 1.98). The readers did not have major difficulties with the texts but found the time available to do the reading the most problematic aspect (see Table 4.15). In 70% of the cases, the readers found the text useful for their needs. In only 6.7% of the cases were the readers not satisfied with their choice of text (see Table 4.16).

Table 4.14: Storing information from the text

Storing Information from The Text		Mean / Std. Dev. (N: 67 min: 1 max: 5)
1	I remembered where the relevant information was in the text.	2.72 / 1.76
2	I made notes on the relevant points from the text as I went along.	2.95 / 1.82
3	I highlighted the parts of the text with a marker (or on-screen highlighting tool) as I read.	2.84 / 1.81
4	I copied and pasted (or copied down) relevant parts of the text as I read.	1.82 / 1.40
5	I transformed ideas from the text into a draft of my assignment as I read.	1.98 / 1.40

1 = Definitely disagree, 5 = Definitely agree

Table 4.15: Difficulties

I found the text difficult to read in terms of...	Mean / Std. Dev. (N: 67 min: 1 max: 5)
the time available to do the reading	2.03 / 1.67
the subject matter	1.41 / 0.96
the vocabulary	1.18 / 0.56
the grammar	1.07 / 0.25
the organisation of the ideas	1.18 / 0.62
the length of the text	1.48 / 1.15

1 = Definitely disagree, 5 = Definitely agree

Table 4.16: Usefulness: The text was useful for my needs

Usefulness	Percent (N: 67)
definitely disagree/ mostly disagree	6.7%
neither agree nor disagree	23.3%
mostly agree/ definitely agree	70%

1 = Definitely disagree, 5 = Definitely agree

4.1.2.6 Discussion of the Diary Study

The diary study was a longitudinal, qualitative study that made it possible to investigate in detail the reading behaviours of a group of students who were preparing assignments for their courses. The main focus of the study was to establish the types of careful reading in academic reading. The readers were reading purposefully for a genuine task and they provided information on the aspects of their reading through a period of time using diary forms. The diary forms elicited information on 'the purposes' with which the readers attempted the texts, i.e. what and how much they wanted to get from the text and 'the cognitive processes' they used, i.e. what skills and strategies they used to attain their goal. The purposes and the processes parts included as many items as could be extracted from the theory and previous research, and helped us to identify the relative importance of and the links between the purposes and the processes in the whole process. This is why the diary forms included not only careful but also expeditious reading operations.

To summarise the information the diary forms elicited on the issues other than reading operations, it is seen that the students participating in the study depended on mostly books, book chapters and journal articles as the sources of information and they did their reading on paper based material rather than online. The greatest source of difficulty reported by these students was the lack of time available for reading. Complicated subject matter did not surface as a source of difficulty for the diary study participants, possibly because they had been reading on the topic sufficiently long enough to familiarise themselves with the subject matter of their dissertations. They were mostly satisfied by their choice of text.

The importance of note taking or highlighting sections of text emerged from the study as important means of storing information for later use. The participants reported that they remembered the information, made notes and highlighted the texts rather than immediately attempting to transfer information to written form. The readers' frequent reports on remembering the information may be indicative of the fact that they were storing information in their long term memory, thus learning from the text. It is suggested in the literature that (i.e. Kintsch and van Dijk, 1983), comprehension involves assimilation of information into previous knowledge and retrieval of it from the reader's long term memory. The readers in the study indicating that they were able to remember what they had read, were apparently storing information in their knowledge base and using it later on.

In general, the reading diary study revealed a predominance of careful reading strategies in the reading operations. All three levels of careful reading were engaged frequently by the readers. However, reading purposes requiring building links across texts appeared as more important purposes among others. Judging the relatedness of texts, evaluating the writer's ideas and comparing viewpoints suggested the importance of reading beyond the limits of a single text and building informational links between texts (documents model).

This is also supported by the fact that the readers reported the predominant use of topic knowledge in understanding the text they were reading. While reading for their dissertation, participants were constantly using their accumulated topic knowledge in making sense of the new text. Understanding a text in its entirety (text level comprehension), i.e. how the main ideas in the text relate to each other, was also an important goal for these readers (2nd most endorsed). For this, the readers reported forming a mental summary as they read by processing a text on a sentence by sentence basis. The conclusion from the general reading purposes and processes was that the readers both attempted to form a unified understanding of the informational content of a text, and to integrate this into the wider context of the topic, necessarily implying building up links between previously read texts.

As expected, this finding was entirely in line with the literature that emphasises that in academic settings where reading is normally integrated with other activities, students are required to construct meaning from multiple sources (Stromso and Braten, 2002). In relation to writing, this is suggested to be *discourse synthesis*; composing a new text by selecting, organising and connecting information across documents (Spivey and

King, 1994). Hartman claims that the process of integrating information from multiple texts requires the readers to 'transpose texts into other texts, absorb one text into another and build a mosaic of intersecting texts' continually constructing a mental web of meaning through the reader's linking of different text-based knowledge sources during reading (1995, 524). Rosenfeld, et al (2001) emphasise that reading activities in academic settings involve reading text materials carefully to remember major ideas for later use, comparing and constructing ideas in a single text and across texts and synthesising ideas in a single text and across texts. These activities require not only content knowledge but also knowledge about discourse patterns, and textual cues, as also marked by the readers in this study who reported that they frequently resorted to their knowledge of organisation of texts.

When fast and slow reading sessions were analysed separately, it was found that what most clearly distinguished fast reading from slow reading in this data set was the group of expeditious strategies pertinent to fast reading such as *1. I used the headings, titles, contents page or index to locate information quickly* (86.5% and 29.6% respectively) and *13. I used my knowledge of how texts like this are organised to find parts to focus on* (67.6% and 29.6%). This suggested that *fast* reading is dual-oriented in the sense that the reader actively uses knowledge of how information is organised in an academic text and quickly assesses its relevance in order to focus on and carefully read only those parts that are deemed to be important. Therefore, in line with its definition, fast reading was *expeditious* in the diary data, punctuated by interludes of slower careful reading at local level. On the other hand, in *slow* reading, the reader follows the writer's presentation of the information in the text and does not distinguish between important and less important information, presumably because the whole text is considered relevant. This is supported by the finding that in reading slowly, the readers marked that they wanted to understand the information in the whole text, judge how it opposes or supports previously encountered information, and evaluate and compare the writer's ideas.

When the reading sessions were *slow*, the readers processed the text, sentence by sentence from the beginning to the end forming a macrostructure of the text. The readers also reported that they combined information from other texts. This is clearly in line with the assumptions of careful reading models emphasising bottom-up processing at initial levels where linguistic input (propositions) is incrementally combined into a coherent mental representation; a text base, which is then integrated with the comprehender's knowledge base; the situation model. It is stated that it is at this last level that not only comprehension but learning takes place (e.g. Kintsch and van Dijk, 1978, Kintsch, 1994, Goldman and Rakestraw, 2000). We have also been able to identify certain careful reading purposes pertaining to *knowledge-transformation* tasks (evaluative, elaborative reading) correlated with processes pertaining to intertextual reading by which the information in the text is combined with the information from other texts read and retained in the memory (Langer and Flihan, 2000; Britt and Sommer, 2004). Therefore, this part of the diary study has provided the evidence that expeditious and careful reading has identifiable dimensions both in terms of the readers' goals and associated cognitive processes.

It is also important to note here that certain features appeared as common in both fast and slow reading. Of the reading purposes, 3. *Find support or opposition to an idea I had in mind* (56.8% and 59.3% respectively) and *13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere.* (51s4% and 77.8% respectively) surfaced as important reading purposes in both data sets. This suggests that in assignment preparation, readers have a specific topic in their minds and continuously compare what they have read up to that point to the incoming information from the new text, judging the relevance and worth of it, no matter whether they read expeditiously or carefully. Relevant to this, in both fast and slow reading, readers depend heavily on *topic knowledge* (*14* - 91.9% and 81.5% respectively), which is again suggestive of the importance of building up links between newly incoming information and the previously learned material.

The correlations found in the diary study between reading purposes and cognitive processes also suggest that expeditious reading involved a complex of processes including selecting relevant parts of a text using topic and textual organisation knowledge and reading carefully only selected parts. Initial encounters with a text seem to involve a mixture of strategic processes. Readers try to locate the information they need as quickly as possible by using their background knowledge of the subject matter and textual organisation cues. They then assess the relevance of the information that appears in the text for their immediate purpose. Expeditious reading episodes are interspersed with occasional careful reading sessions in which the readers focus on parts of the text they find relevant and important. On the other hand, the purposes associated with careful reading (8. *Understand and remember as much information as possible from the text as a whole*, 10. *Form an understanding of how the main ideas in the text relate to each other*; 11. *Evaluate the writer's ideas on the topic*; 12. *Compare the writer's point of view with that of the other writers*; 13. *Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere*) are closely related to the processes of reading incrementally from the beginning to the end, extracting information from sentences, rereading and progressively building up a coherent macrostructure of the text in mind (situation model) and through building up intertextual semantic links, forming a documents model.

Therefore, through the diary study three main points have been identified. Firstly, both expeditious and careful reading operations are used by the readers in assignment preparation, the latter being predominant.

Secondly, although correlations in a small data set such as this are not sufficient for conclusive statements, it can be claimed that the links found between the reading purposes and the processes are in line with the proposed definitions of both expeditious and careful reading in theory and helped to confirm that they served different functions as suggested in the literature (Kintsch and van Dijk, 1978, Kintsch, 1994; Urquhart and Weir, 1998; Khalifa and Weir, 2009). Thirdly, careful reading appeared to have three identifiable levels; *sentence*, *text and multiple-text* levels, the last of which appeared as the most crucial skill in preparing assignments.

4.1.3 Interview Study

The six participants in the study were each interviewed once, after submitting their diary forms. The interviews provided an opportunity to follow up with the participants the issues that had emerged from the reading diaries and to carry out an additional, supervised reading session in which the participants could comment in concurrent and retrospective think-aloud procedures on the process of reading as they engaged in it.

The data are presented below according to *'reading purpose'* categories listed in the diary form (Section 4.1t2 and Appendix 3.2). Since the interview data were meant to provide confirmatory (or non-confirmatory) evidence to the nature of reading purposes and the associated cognitive processes as they were defined in the theory, and as the data came from a small sample, no statistical analysis was done on them; instead, illustrative examples of how the readers set up their reading goals and how they acted them out – i.e. using which cognitive processes – are reported here.

Moreover, the interview data yielded other important strategies that were not covered by the reading diary form, yet pertinent to reading across texts. The data confirmed that, as put by Rouet (2006), documents reading included – together with the representation of the situation within a text – information about how different texts relate to each other (Braten et al., 2009; Goldman, 2004; Perfetti et al 1999). Content evaluation was in fact an important skill captured through the items 11-13 in the diary form (shortly; 11. evaluate the writer's ideas, 12. compare writer's ideas, 13. judge how this text was related to other texts). However, the evaluation of *content as a source characteristic* is taken as 'content evaluation based on the reader's own opinion and the elaboration provided in the text to support arguments (see 3.1.3.2).

This study also revealed that source characteristics were taken into consideration when the readers formed an understanding of the situation described in the text and how that description of the situation compared to what they had read about the same issue before. In the current analysis, four categories of source characteristics, *author, document type, content* and *date* are observed and they are given in Section 4.1.3.2 below.

More importantly, the interview data revealed another significant evaluation strategy that went beyond evaluating mere content relations between texts (how two texts were related) and involved evaluating the content in relation to the texts that the readers were already in the process of developing themselves; *usability.* The over-arching criterion for the inclusion of information in the emerging text was its direct relevance to the readers' dissertation. For the information to be included in the text, it had to be relevant to a specific part of their text; fitting in a part of the text so that it could be deemed *usable. Usability* of the information in the text the readers were reading in relation to the texts they were writing (their dissertations) was an important criterion.

Unlike in the previous studies that investigated documents model reading in rather experimental mode (ex. Hartman, 1995; Stahl et al, 1996; Wolfe and Goldman, 2005), in the present study the readers had authentically been involved in reading for an extended period of time in the field of their interest and they had been developing a text of their own on a specified topic. The documents they read involved both relevant and irrelevant information for their purposes and they had to cull through a considerable amount of information to find what was usable in their own text. Although the texts they read were normally in the general area of their interest, they could differ from each other in several aspects such as focus (i.e. the specific research question addressed), methodology, subjects, context, and theoretical orientation. The readers needed to evaluate how and to what degree the content of the text they were reading was relevant to their dissertation; if relevant, specifically at which parts of the text, and whether the information was usable in their own text; if usable, in which part of their dissertation.

The evaluation of *usability* of information was an over-arching strategy for the readers reading across texts with the purpose of fulfilling a writing task. This strategy shaped the reading processes of the readers significantly as they needed to not only form a situation model of the text they were reading and compare it to what they had read before – therefore, form a documents model across the texts – they also had to evaluate to 'what extent' and 'from which aspects' the situation given in the new text was overlapping with their own task.

This was actually building a many-fold documents model across the texts read and between these texts and the text being written. This is evident in the literature that suggests that the reader's purpose or goal in reading is an important determiner of the kind of representation that will be formed (Goldman, 1997) and such higher level tasks as discourse synthesis, in this case the construction of the dissertation, require organising of content by supplying a new organisational structure to the information gathered through various texts – which may not be explicitly linked – and connecting it by providing links between related ideas and synthesising the information to produce a new text (Spivey and King, 1989; 1994). In judging the usability, the readers had several secondary criteria. One such criterion was *novelty;* whether the information was new, something the reader had not encountered before or something that brought in a new interpretation to the situation and therefore worth using.

On the other hand, in certain cases, the re-occurrence of certain information in more than one document confirmed the significance of it, thereby making it 'more reliable' and apt for inclusion in their dissertation. This criterion is named as *cross-textual support* in the categorisation. Another important criterion the readers had in relation to relevance/usability were the *detail* the author provided in the discussion and the *clarity* with which the arguments made. These characteristics, together with source characteristics, were used in choosing the information to be focused on for a coherent, detailed and critical evaluation of the situation across documents and selecting information for forming of a unified and a novel discussion of the situation. The judgements of source characteristics/trustworthiness and relevance/usability are discussed in Section 4.1.3.2 and 4.1.3.3 respectively below.

Another note-worthy finding from the interviews was the strategies related to how the readers made use of the texts beyond content comprehension. In academic settings, reading multiple documents usually entails an end product either in written or oral form. Understanding and learning the content of texts is the prerequisite for such activities as assignment preparation. However, it is a known fact that deep processing of texts does not only end in learning the content but it also facilitates the acquisition of linguistic and textual knowledge; acquisition of vocabulary and syntax and genre-dependant organisation of information in the text (textual organisation). In this study as well, reading was observed to be instrumental in learning both language of argumentation and textual organisation for research studies. The readers in several cases reported that they sometimes read to learn how the writers of the articles organised and presented their arguments, how they used the language to argue, they compared their own ways of selecting and paraphrasing information with those of the writers who they deemed have more authority. The readers also read to learn about methods of research and how their study compared to similar studies in terms of quality of research. Research articles provided the readers templates to learn how to use the language for argumentation, how to organise the argument structure of their texts and how to amend the methods of research in their study. These are exemplified in Section 4.1.3.4 under the heading '*Learning beyond content*'.

4.1.3.1 Reading Purposes

As explained above, the interview data are categorised and presented here according to 'reading purposes' in the diary form (Section 4.1t2 and Appendix 3.2).

1. Find specific words, phrases or figures: Specific words seem to be relevant at the stage where the participants looked for relevant articles on the web. When they were searching for articles, they used a key word search strategy via search engines or individual journal web sites.

'I searched for the key words on the internet and then I went into journal websites as well'.

At the initial stage, they also looked at the context in which the key words appeared in the text.

'There were some key words that were relevant to my dissertation. I briefly read the sentences in which they appeared'.

2. Locate the parts of the text that would be worth reading later: The participants often read through their articles quickly, assessing which parts of the text were relevant for their needs. Numerous comments on relevance are found in the data. One participant, in an illustrative example, explained that when she went back to an article, she only read the highlighted sections but did not read the rest. For her, highlighting parts of a text meant locating the parts that she would read again.

'Now, I am highlighting this part. I will go back to this part and read only that part later on.à

'So what I do is to look at a few sentences from each paragraph and go through very quickly looking for some words and understand whether it is talking about something that is relevant. I read the first part of the sentences and skip explanations, examples etc. I can have a rough idea

what the text is about. The second paragraph is not very relevant because it talks about think aloud procedure.ś

'I have just skimmed this paragraph. I decide that some part is important when I see important words such as 'decision style'. Only then I read carefully.ś

3. Find support or opposition to an idea I had in mind: The participants who had already begun to form a written text would mainly look for support or opposition relevant to the arguments they had raised.

'I want to find more information to add on the literature review part where I discuss reading strategies, previous research that has been done.ś

'I would like to support this statement saying that students with higher abilities in English, they use more strategies, more combination of skills, etc. when they respond to a particular item.ś

'So what I am going to look for is something new, something different from what I have read. Or something that can support, related studies which are new and which have findings that can support my arguments.ś

'I am reading on this subject just to find more support for these findings. My second research question is on this topic.ś

4. Find specific information to answer a question: Nothing of relevance to this purpose emerged from the interviews. This is probably because of the nature of the assignment that these students were working towards. They were not answering a question, as in many student assignments, but were working on a self-selected research project. For the think-aloud session, they had chosen an article they believed would be relevant to their area of research and were trying to establish what might be useful for them. The participants certainly had their focal topic in mind as they read, perhaps with research questions to be answered, but were not looking for a specific piece of information to provide 'an answer'. Rather they seemed to be dealing with several, sometimes not clearly defined questions at the same time.

5. Get a basic understanding of just the main idea(s) in the text: The participants were able to form the gist of the article in their minds by quickly looking through the pages to find key words, reading the introduction and the beginnings of the paragraphs. This appeared to serve the purpose of helping them to decide whether the text or parts of the text were worth reading carefully.

'By quickly looking through the pages of this article, I can say that this is a study done in Taiwan. It is not only theoretical review but it includes some research findings and statistics.'

'I am going to read the introduction. This is just giving me some idea on the background information on skills and strategies.'

'I read the introduction quickly just to get a general idea.'

6. Decide whether the text would be worth reading: As mentioned above, the abstract of an article was of crucial importance for these participants in deciding whether the text would be worth reading.

'I chose this article on the basis of relevance of the topic. Just looking at the abstract was enough for me to choose this article.'

'First I want to check what it is about. I will go to the abstract to see what is going on.'

'How successful can a pre-sessional course be? I am interested in that because I teach at pre-sessional courses. I do take my time to read the abstract.'

Looking at the headings and reading a few lines of each paragraph was also important.

'I clicked on the article and ran through the abstract. I read the whole abstract carefully. Some research questions seemed to be related to mine. Not all but some of them are relevant and some of them will be of interest to me later. Then I looked at subheadings, looked at what each heading was discussing quickly. And I have decided I will read it in detail.'

'I read the abstract carefully, looked at the introduction quickly to set the general topic in my mind and I looked at the first lines of the paragraphs. I decided that I want to learn about different approaches to writing because it is closely related with my dissertation topic in mind.'

'I immediately read the abstract on the webpage. I normally look through the pages and look at the subtitles, read a few lines from each paragraph quickly to decide whether all the paragraphs are worth reading.'

'I have no time to read everything. So what I do is to look at a few sentences from each paragraph and go through the text very quickly looking for some words and understand whether it is talking about something that is relevant. I read the first part of the sentences and skip explanations, examples etc. I can have a rough idea what the text is about.'

'All the key words in this article are related to my study. I think the whole passage is very relevant.'

'I have just skimmed this paragraph. I have decided that some part is important when I see important words such as "decision style". Only then I read carefully'.

Scanning for key words also helped some participants to decide whether the text is appropriate to their purposes.

'I can't find any key words here. I wouldn't really read this article. It is not specific enough.'

7. Check my comprehension of a specific part of the text: Since these participants were reading the articles for the first time, they did not reread a specific part of the text to see whether they had understood it correctly. However, one participant did need to read some parts of her text twice as she lost her concentration at points.

8. Understand and remember as much information as possible from the text as a whole: The participants who found relevant and new information in a text did then read the whole text from the beginning to the end. This appeared to serve the purpose of learning about the topic. This was most apparent in one of the participants who had just started to read about approaches to teaching writing; she was reading about something she had heard but had not read about before. She read the whole text from the beginning to the end without skipping any part, building an understanding and developing expectations from the rest of the text as she read along.

'Now I am moving on the process approach and I am wondering what it is going to be. I see, OK. It is interesting... What is interesting is that we are starting with the point of view of the students and going towards the product, not vice versa. (reads)... less importance on structure... but structure is important (reads). What I do now, I constantly refer to what I read initially, to product approach so I can compare easily both approaches.à

'If an article is very interesting, I find it quite difficult to focus on what I exactly need to focus on for my assignment. I read the whole article. That's why I sometimes waste time. When I am interested, I read to learn.à

9. Understand and remember as much information as possible, but only from relevant parts of the text: Those participants who had already accumulated considerable information on the topic they were reading about seemed quite efficient at deciding on the parts they should read carefully. Since the text they had in front of them was a relatively short article, not a lengthy book, their primary aim was not to locate the parts that would be worth reading later, but to focus attention selectively on the parts that would be useful for them as they went through the whole text.

‘This part is on the reading strategies of successful and unsuccessful readers. This is very directly related to what I am doing now. So I will read on.ś

‘Metacognitive awareness is not relevant to my topic so I am not going to concentrate on it. I am going to focus on reading proficiency and reading strategies so I am skipping this part.ś

‘This study is not very relevant for me. It is conducted among children. I will quickly have a look but I am not going to read in detail.ś

10. Form an understanding of how the main ideas in the text relate to each other: The purpose was strongly endorsed especially when the participants read to learn and formed a summary of the whole text in their minds, for example the participant reading to learn about the approaches to teaching writing. Two other participants read the whole text, however, mentioning some parts of the text was not directly relevant for their own study. They disregarded the less relevant sections when they summarised the information from the text that they deemed important. In some cases, particularly when participants identified that a reference given in the article they were reading would be a key text, they made it clear that they would read the referenced text carefully as a primary source.

'If an article is very interesting, I find it quite di t I exactly need to focus on for my assignment. I read the whole article.è

'I like to get the best out of what theories off.. So I am .ry excited about learning these approaches. That's why I want to read this article.e... Now, I have learned product and process approaches as well.

'I read this article from the beginning to the end because there were several important points. Some of the parts were less useful though.è

I have read everything up to this point and I know what the writer means. But I will skip this part where he writes about children.'

11. Evaluate the writer's ideas on the topic: Critical evaluation was an important focus in the reading processes of the participants. The readers frequently commented on content and the quality of the ideas, i.e. how the arguments were supported in the text and whether they were congruent with the reader's background knowledge. Critical evaluation could be done within a text span depending on the world knowledge of the readers (as in the first three examples below) or it could be done in relation to the information from other texts (as in the last example). Since evaluation and elaboration of the information in a text is an integral part of academic reading and it is a many-fold complex process, many samples of interview protocols include evaluative comments that can be categorised in other categories as well (ex: comparison of writers' points of view, source evaluation and relevance/usability).

'When I read the product approach, I found it had a lot to offer. But when I read about process approach, I saw that it could be complemented with strategies from process approach. But that didn't put into question the validity of what I read before. I don't believe in one theory as opposed to another one or another one. I believe every approach has something to offer.'

'Hum, an interesting phrase, 'assisted imitation'... First you see how it is done, then imitate. Then there will be more freedom you can actually use what you have learned. There should be a next step where there would be more freedom. Ah, it is there... It says in the free writing stage students should feel as if they are creating something of their own. Perfect!'

'This author thinks that these students didn't acquire sufficient English and this is why they are having language related problems. I agree from language point of view. When students come to the UK they have language problems but they also have cultural adjustment problems. This author thinks that the main challenge for the students here is the English language. I think cultural problems are very important, too.'

'I have read something here and it rings a bell. I have just thought that ESL students were not good in their performance in English but may be they were not good in their first language either. This was mentioned in an article I read previously but in the present article, it is not taken care of. I think this should have been accounted for.'

12. Compare the writer's point of view with that of the other writers:

One of the primary goals of a dissertation is to bring together congruent and contrary arguments. Thus, the participants needed to find supporting or opposing information for the arguments they needed to discuss in the body of their research. Therefore, frequent comments relating to this reading purpose existed in the data.

'I want to learn about the author's view based on his findings and also to look into what other researchers have said about this.à 'In research articles, there is usually a debate, just like here. I pay special attention to the points of arguments because I want to know what the opposite point of view is, what they argue for or against...'

'Here, he talks about typical problems in daily life. These problems are pretty much the same with what other scholars mention.à

'I want to learn about the author's view based on his findings and also to look into what other researchers have said about this.à

'There are two more sentences here that are important. They summarise the findings of another study. Surprisingly, it says here that both high and low scoring students appear to use same strategies. This is a bit contradictory to what I have read earlier.à

13. Judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere: This was perhaps the reading purpose that drew most comments in the think-aloud sessions. Judging the relation between texts surfaced as a crucial strategy with which the readers compared the information across texts and decided how each bit of information was related to others. In order to do that, they needed to establish intertextual semantic links between the texts and evaluate whether the texts were corroborative, contradictory, if so in what ways, or whether they filled in the gaps in others by providing the missing information complementarily, whether one updates a situation in another by recent findings, etc. The referencing across texts was also paid attention to. In some respects, judging the relation between the texts necessarily overlapped with other categories such as comparison of writers' ideas, however, it involved establishing more varied intertextual semantic links (i.e. contradiction) some of which are cited above: As these readers were reading research articles, they placed the research study in a context of research focus, methodology, subjects in the study, the date and the place of the study, and relative to these, they evaluated to what extent the information in the present text compared to other texts.

'Lao (not mentioned in the article) also made a comparison in English and Chinese and reached the same conclusion. Better participants always use more strategies.i'

'I remember Paris and Jacobs. They said the same thing that first language learners, who are actually more proficient as compared to ESL students and poorer L1 participants use more strategies.'

. ..now that it i. Swales who origina , defi ed genre as a c ss of communicative event the members of which share a set of communicative purposes. These writers use the same approach.i'

‘Anderson, a researcher I read before, investigated differences in strategy use on two reading tasks. The methodology is not exactly the same as here but he has also referred to meta-cognitive strategies and reached similar conclusions.’

‘Since this author is a visiting lecturer, he had the opportunity to talk with some students studying here. Also, his position can provide him access to true information and through his teaching position back in China. It is worth comparing this type of study with the others.’

‘In the introduction, the writer mentions an important researcher. I use his theory as well so it is very important for me to see what kind of information this writer used in this article, how this writer made use of the theory.’

‘I found the reference of this article in another article I had read before. There was a quotation from this article in the previous one. From that quotation, I could imagine what sort of a research study this article related.’

‘There is more on individualistic society here, a good comparison. This definition is more extensive than the other definitions I have read before because the writer explains it from an intercultural point of view.’

4.1.3.2 Source Characteristics/Trustworthiness

As explained before, source characteristics surfaced as an important strategy in documents model reading. The readers, in judging how different texts were related, used information about the sources themselves and what the characteristics of those sources were. The content of the text was only assigned importance when certain source characteristics were deemed positive. That is to say, the document type, for example, was an important criterion for the readers to regard the information in the text as 'reliable' and thus, worthy of use. If a text appeared for example in a recognised journal or recommended by an expert, the text was considered 'trustworthy', and the information in it was considered reliable and usable. As mentioned before, 'source characteristics' appear closely linked with 'relevance/usability' criteria in this study, and among the 'source characteristics listed by Rouet (2006) and Braten, et al. (2009), *author*, *document type*, *content* and *date* surfaced as important qualities in judging the value of a document in reading, selecting and combining information from multiple texts. An exemplary list of comments is given below grouped according to the source characteristics mentioned in the literature. However, some participant comments carry information relating to more than one category and they appear under more than one group heading.

Author: The comments on the author were mostly related with his/her authority in the field; whether he or she is a well-known researcher and whether his or her work has been cited in several other sources. Recommendation by a reliable person and cross-referencing in a dependable article were also important in deeming the author of an article as trustworthy. If the author, thus the source text was deemed dependable, any cross-references were noted as potentially important and the readers mentioned their intention of reading those texts as well:

'In fact, Gardner is the founder of the motivation theory in second language learning. My supervisor suggested him to start with.à

'I refer to Dörnyei a lot. He has published many articles and they are key references for me.à

'Anderson's study is important because it is mentioned in a few articles. I will put all these studies together and then come to a conclusion.à

'I see Paris and Jacobs again. It is all related to my work. I think Paris and Jacobs is a key article. I should read the original work.à

'I will look for other authors who are quoted in this article. This is written by a well-known researcher and it is a key article. The references she cites must be important. I may need to read some of them later on.à

'This book was not recommended to me but I was looking at the shelves and these two people worked with me (writers) in EAP courses. They were very effective teachers. So I thought there might be interesting stuff about EAP in this book.à

'I found the reference of this article in another article I had read before. There was a quotation from this article in the previous one. I looked for the article.à

Document type: For these readers, if a text was from a dependable source such as a refereed journal article, the information was more readily accepted or rejected otherwise. Genre related characteristics the texts exhibited were taken into consideration. The readers evaluated and valued the information from texts differently when a text was, for example a conventional research article having literature review, explicit accounts on methodology and data analysis, or an internet blog; i.e. the method the information was gathered and presented in the text was an important issue. The distinction in this study was between a research article and others – there was not a distinction between *primary* and *secondary sources* as in the studies on historyl– and the main concern was whether the document followed the expected research conventions in Applied Linguistics studies. As experienced readers, the readers exhibited keen awareness on the document type.

'Reference comes from a dependable literature, a dependable source!'
'This second article is more relevant to my research. And it is longitudinal and research based. The author talks about research method, the research instruments used, provides data and statistics. It has a clear structure and you can read the research questions, etc.
'It is a piece of research work and I feel it is reliable.l'
'This author just gives general ideas what the problems are and categorises them. For me, this doesn't tell you exactly what the learning problems are. And he didn't do any research on it.'
'I don't think this article appeared in a journal. If it is just an article on somebody's blog or diary, even if the information in it is true, I don't think I will use it. And in this article, there are not any references at all. It is totally

based on this visiting lecturer's past experience and communication with the students.’

Content: The evaluations of *content* in relation to *source characteristics* were based on both *the reader's own opinion* on the content of the document, and the degree of textual *elaboration* supporting the information in the text.

Content evaluation in terms of how much the reader could trust the document was closely linked with other criteria, for example document type (whether it is a conventional research article), and the evaluation of the information in the text in comparison to previously acquired knowledge. However, in some cases, the readers evaluated the content of the document merely based on their own 'world knowledge' and decided whether the information was worthy of inclusion in their own document.

'And this is a speculation, an assumption… not based on research.’

'I have never seen people categorise students like this but I also know that it is just a fact; some students are simply more hard-working than others. Since it does not have any scientific basis, I don't think it means anything.'

'They (Chinese students) can help their classmates to solve some problems. This last bit may be a contribution but the rest, I don't think it is a contribution.'

'I know from my experience that teaching strategies to students make the process (teaching process) less painful. When I read the product approach, I find it had a lot to offer. I think this part of the article is quite accurate.ó

On the other hand, the readers, looking for information to expand their already existing knowledge, frequently required relevant but also detailed information from the text. It was important for the readers that the information in the text was elaborated enough, that is supported by adequate explanation and argumentation – that is it has sufficient persuasive powerø- so that it could be regarded as trustworthy.

'The first article I read on this subject was too general. This is in more detail. I like this one.ó

'I have read some articles before about learning problems but pretty much in detail. For example, their problems with listening comprehension, or problems with taking notes in the lecture, very detailed. This author just gives general ideas on what the problems are and categorises them. For me, this doesn't tell you exactly what the learning problems are.'

'I think this writer's approach is quite superficial.ó

'This (age difference) was mentioned in an article I read previously but in the present article, it is not taken care of. I think this should have been accounted for in detail.ó

Date: In this study, *date* meant the *publication date* of the document. When the readers judged whether the information in the text was trustworthy, they paid attention to the publication date of the document to ensure the information in the article was not outdated. This is slightly different in comparison to the studies done in the field of history in that the readers did not want to locate the document on a time line to relate it other documents but they paid attention to the recency of the information in it.

'But Gardner is a bit outdated now. Research is becoming outdated very fast.t'

'It is important to read new research articles because you can find many important and relevant research studies cited in them. They usually summarise the previous ones. You can go back the old ones but it is important to know the new studies. Some new findings may have changed the situation.'

4.1.3.3 Usability

As explained above, the reading processes of the participants in this study were geared at finding information to be included in their dissertations. Thus, the main concern of the readers was whether the text provided relevant information to the topic in their mind. That is to say, the readers needed to build semantic connections between the pieces of information. This connection might be of any kind; corroborative, contradicting, complementary or in the form of referencing. If the information was deemed relevant and from a trustworthy document, the readers evaluated it from the aspect of its usability, that is to say, whether the information could be functional in developing their arguments. The concerns on the context of the study— *setting* as Perfetti et al (1999) put it— is included in the category of *usability* here as the evaluative of the context in this study mostly referred to the educational and cultural context of the research study in the articles the readers were reading and once again, the more contextually similar the study they were reading to their own study, the more relevant and thus usable the information was.

All in all, these readers were looking for information to be used in their dissertation and they made frequent comments on *usability*. They evaluated the information as being usable when they found it directly related, when they could see the information could fit in a part of their dissertation and also when they found it *novel*, (*novelty:* if the information has or has not been encountered before), supported across texts (*cross-textual support:* the information appeared in more than one source) and clear (*clarity:* whether the information is presented in a clear and understandable language). Illustrative examples are in order:

'Also I can use some of this knowledge in my literature review. I can use this part to talk about Chinese students' learning background. This may help the reader to understand why a Chinese student may behave differently from western students or other international students.'

‘There are also a lot of definitions and terms here with examples. I think this will be useful to me when I need to put these into my thesis. So I am underlining them. May be I can compare different definitions.è

‘I will just check findings because those parts mean when, why and how. I can include information from those parts in my report.è

‘Many researchers also mention that Taiwanese culture is collective and this supports my point of view as well. I can use it in my literature review.è

‘I can use this piece of information to explain the problems I have observed between Taiwanese school managers and teachers.è

‘I won’t focus on one single approach in my work so I can draw these (product and process oriented approaches) together.è

‘This study is not very relevant for me. It is conducted among children. They read a story and talk about it. Mine is more academic reading.’

Novelty: As the readers had been reading on their topic of interest for a while, they had considerable knowledge on the topic. Given that the information appeared in a trustworthy document and relevant to their topic, they valued new pieces of information as usable and discarded familiar ones.

‘I want to find more information to add on the literature review part where I discuss reading strategies, previous research that has been done. Let’s see if there is anything new?’

‘For me this is not new information. It is a fact, I myself know. Everybody knows this as a fact. It is sort of a common knowledge that if a student

comes here to study, he must have a good financial background. I will skip.’

‘When I read through, I am not sure whether the author has contributed anything to the issue or not.’

‘Here, he talks about typical problems in daily life. These problems are pretty much the same with what other scholars mention. This is again gathered through the interaction with these students. Nothing really new to me, not useful.’

‘I have read about these issues in so many articles. So what I am going to look for is something new, something different from what I have read.’

‘This is something totally new for me and I can use this piece of information to explain the problems I have observed between Taiwanese school managers and teachers.’

Cross-textual support: Another important feature that the readers took into consideration when attending the information in the text was whether the information had been encountered before in other reliable sources. When reading on an academic subject, the selection of information for the assignment was facilitated by the confirmation of its importance through its reoccurrences across texts. In scholarly articles, ‘referencing’ is a frequently used technique and it attests the trustworthiness of a document. However, cross-textual support here refers to the pieces of information rather than the whole source document and it is taken as a confirmatory factor for the importance, therefore for the inclusion of it in the target text.

Obviously, seeing the similarity of information across texts (covered by the 13th reading purpose above) is a pre-requisite for this.

'Now they talk about decision style. They say it is relative to culture. I have read similar discussions before. It should be an important issue in intercultural communication area.è

'That was mentioned somewhere else too. I think it is important and I will put it down in my research. I would like to support this statement saying that students with higher abilities in English, they use more strategies, more combination of skills, etc. when they respond to a particular item. So this supports my assumption and I should take note of it.ė

'Strategies are used to solve problemsè This is something I have read earlier. I think this is a general belief. I should mention this too where I am discussing the issue.è

'He (the writer) mentions somebody else here. That researcher also supports his point of view. I haven't read this reference, though. May be I should read.ė

'I wouldn't say that I have learned tremendously new things but it has reinforced what I already knew. At the initial stage of the study it is good to see certain things are reinforced. So you know you have to mention them. It may be interesting to see an article totally against the genre approach at some stage, though.è

Clarity: Although quite few, some comments appeared on the clarity of language in the text. Apparently, the readers were more able to extract information from the texts whose language and sophistication level matched theirs.

'As I read along, I think this is a clear text, well explained, goes to the basics, which is what I really need at this stage. I am pleased that it explains concepts clearly. I can use these explanations.è

'I like the way this second article is written. It is easy, clear, reader-friendly reading this and it flows smoothly. This kind of article is more helpful.è

4.1.3.4 Learning beyond the Content

Another observation made through the interviews was related to how the readers made use of the texts other than looking for relevant information to integrate in their assignment. The readers in several cases reported that they sometimes read to learn how the writers of the articles organise and present their arguments, how they use the language to argue, compare their own ways of choosing and paraphrasing information with those of the writers who they deem have more authority. The readers also read to learn about methods of research and how their study compares to similar studies in terms of quality of research. Research articles provided the readers templates from which they could evaluate not only their use of content, but also their language use, text and argument structure and methods of research. Here are a few examples:

'When I read methodology in detail, I would like to see what methods others have used and whether I can make use of them or not. Instruments, statistical analyses, etc.i … But I use a proforma, he does interviews. I wonder whether I should do interviews, too. Reading all these studies may give me ideas about how to improve my own study, too.i'

'In the introduction, the writer mentions an important researcher. I use his theory as well so it is very important for me to see what kind of information this writer used in this article, how this writer made use of the theory. I want to look at how he describes the theory. He is using the theory as his starting point as well. So this article is very important for me especially the literature review.i'

'In research articles, there is usually a debate, just like here. I pay special attention to the points of arguments because I want to know what the opposite point of view is, what they argue for or against and how they do it .. so that I can adapt their way of argumentation.i'

'Here he mentions Hofstede 1994. I actually read his book so this is information I already know actually. But I can see how people quote his studies. I pay special attention to how writers paraphrase other writers I have already read. I compare the original texts with the paraphrases. Even for the same quotation, different writers use different phrasing. Everybody's styles are different so for me it is a way to learn how people quote stuff.i'

4.1.3.5 Note-taking

In the interviews, this seemed to be an important intermediate step for the participants between locating or selecting information and integrating it into an assignment; the participants did not transfer information directly into their emergent texts but highlighted the text, which may simply be due to the fact that they read in the researcher's office rather than their own study environment as one participant said she would type it on a computer in a file where she collected useful information.

'I like to mark the text with a highlighter if it is a key article. If I can't do that I feel frustrated. I usually start reading with my highlighter and I mark a lot.'

'I usually make notes mentally. I never take notes at the first time. I read and highlight and I have to read again when I need to write. When I go back to an article for the second time, I tend to focus on the highlighted bits.'

'I will highlight this part. It elaborates on 'individualism'. If I were home, I would type it in. I prefer to collect similar information in a file and when I am writing up, I select the most useful ones.ò

4.1.3.6 Discussion of the Interview Study

The interview study reported above enabled us to have a closer look at the reading operations used in an academic setting. The participants were reading for a genuine purpose and they were dealing with a complex academic task that required the highest level of literacy skills. The six postgraduate students involved in the diary study attest to the importance of reading critically at both the text and multi-text levels when preparing an assignment. The participants compared information in the texts they were reading and evaluated it in the light of what they had previously read. In this process, their considerably firm topic knowledge was of crucial importance and facilitative (Spivey and King, 1989; McNamara et al 2007; Perfetti et al 1995). According to purpose, the participants made use of expeditious reading strategies to locate required information and used careful reading for detailed comprehension of selected material. Expeditious reading for the participants often involved a focus on a particular section of a text with the aim of locating relevant information. Background knowledge, both of the topic and of features of textual organisation, was important and the participants usually knew what they were looking for when they approached the text. Episodes of careful reading at local level appear to have formed an integral part of expeditious reading. Often slow, careful reading appeared to be selected when the material was unfamiliar or had been identified as being of central importance to the assignment topic. When the readers were not interested in forming a representation of the whole text and wanted to locate a particular detail to focus on, they read expeditiously and skipped irrelevant parts. In these cases, slow incremental reading, after the selection of the relevant material, may have helped the participants to form a coherent cross-textual mental model of the predetermined problem in their mind.

Careful reading episodes seemed to be related with retention of information in long term memory and were clearly intertextual.

The interviews made it clear that academic reading is a multi-layered activity in which searching for, understanding, evaluating, selecting, organising and connecting information required different cognitive processes actively orchestrated by the reader as needed. The literature strongly suggests that academic tasks are purposeful activities that require skilful knowledge management in which not only bits of information are understood in isolation but all bits of relevant information are meaningfully and systematically related to one another (Goldman, 2004) resulting in production of new knowledge from already existing one; *knowledge transformation* (Bereiter and Scardamalia, 1987), which is constructing a single, integrated model of all the situations described by various authors (Britt and Sommers, 2004). As stated above, this requires explicit criteria for the selection of information that will provide evidence for the reader/writers' arguments in line with their goals, elaboration of information by connecting it with prior knowledge, extensive planning and monitoring (Spivey and King, 1989). The reader needs to construct new types of intertextual connections taking into consideration disagreement, discrepancy and other such qualities of the information and form a mental representation of a 'tentative truth' as they read along. Hartman (1995, 525) points out that establishment of such links, i.e. intertextuality, is placed in *discursive habitats* (discourse stand) and in academic settings, 'the convention-governed habitats of the research process ... defines the links that can (and cannot) be made.' Goldman and Bloome (2004, 5) also determine that each discipline has rules that govern how multiple texts are related to one another and what the integrated representation represents. In line with these suggestions from the literature, it can be seen in the interview data that the careful reading processes of the participants in this study were strongly intertextual as even though they were attending only one article they brought to the study, they were comparing and contrasting information in the text at hand and what they had read before, and actively building new semantic links between the pieces of information (ex: *x* mentioned here opposes *y*, which I read in a previous article), evaluating their relative standing against each other (ex: *x* and *y* are on the same issue of *A*) and creating an entirely new situation that was non-existent and beyond what was described in the texts they read (ex: then, *y* cannot be taken for granted without taking *x* into consideration). It could be observed that they had clear ideas what they could and should include in their dissertation (ex: I should compare and contrast *x* and *y* in my essay) and were organising and planning information for their text along with the conventions of research as they dynamically created a mental model of the situation across texts (ex: I should mention this discrepancy when I talk about *A* in my literature review chapter). The participants evidently needed content knowledge, knowledge about text and knowledge of relationships between texts in order to relate ideas during reading comprehension (Perfetti et al, 1995). The emergent result of performing such a task of reading across multiple documents to form a written text evidently required the actual construction of an evolving documents model in which several texts were connected and situations described in each text integrated in a higher order situations model (Rouet, et al, 1996; Perfetti, et al, 1999, Wineburg, 1991).

Interview data also confirmed evaluation of source characteristics as an essential component of the documents model. As stated above, Perfetti et al (1995) stress that connecting information at the documents model level requires motivated strategy use, extensive content knowledge, knowledge about the text itself as well as knowledge of relationships among texts. Rouet (2006) sees source evaluation as an integral part of documents model reading as it is source information that helps the reader to differentiate documents and evaluate respective contribution of each document to the global representation of the situation. The readers in this study could critically evaluate the texts against each other making comments on such source characteristics as *author*, *document type*, *content* and *date*. The characteristics involved in these categories changed in accordance with the discipline of Applied Linguistics as opposed to History. Document type, for example, did not involve considerations of *primary* versus *secondary sources* as in History (Perfetti et al, 1995; Wiley and Ross, 1996, Britt and Anglikas, 2002), but it entailed *conventional research articles* versus other types of texts. Evaluations on content not only involved the reader's opinion but also considerations on the amount of textual elaboration the article provided to support the arguments. This is an important finding as textual elaboration is seen as functional in facilitating the reader's updating situation model upon encountering new information (van Oostendorp, 1996). Date gained slightly different meaning denoting publishing date rather than historic period or context. This provided us with the preliminary ideas as to what source characteristics may be attended in Applied Linguistics since the study above is the first one done in the field.

Another note-worthy finding was explicit comments on how the readers could use the information from the document in their assignments (usability). They had their embryonic dissertations in mind as the destination for the new information they were accessing in their reading and this was generally their basis for selecting and evaluating the material. Participants appeared to be building a mental model of their assignment as they read and considered where new information would best fit in to their developing plans. The information to be selected should be not only relevant but also usable in the context of their text by bringing in recent, confirmed and clearly expressed knowledge. These processes reflect a complex planning behaviour involving selection of right type of information and its purposeful integration in a coherent whole. This is an integral part of reading multiple documents for a writing task (Spivey and King, 1989).

Document level reading for the purpose of writing a research study also enabled readers to acquire certain conventional aspects of research studies. As has been exemplified through the comments of the readers, the readers, through an on-going process of reading, were improving not only their topic knowledge but also their knowledge base on text structure including specific characteristics of certain genres, their linguistic skills as well as their writing potential. Therefore, they were learning from the sources how the meaning is constructed and represented in the field of Applied Linguistics.

To sum up, the interview study made it clear that readers reading for writing purposes construct a clearly defined goal. Reading processes are affected by the requirements of the task and readers strategically adjust their reading process according to the requirements of the task (Pressley and Afflerbach, 1995; Braten and Stromso, 2003; Gil et al, 2010). Reading for the completion of a task, as is normally the case in higher education settings, is across-texts by its nature. Readers who have accumulated content knowledge know what they need to get from the text and they are sensible to the different functional value of the types of documents for constructing an argument although the criteria may change across disciplines (Rouet et al, 1996). They evaluate the sources critically and they have explicit criteria for selecting information to be included in a written assignment. Documents form templates for readers to learn discipline related conventions as well as linguistic features. In reading across texts for writing an assignment, readers form a documents model that combines the situations across texts and they place each text in a context depending on its source characteristics. All these are in turn linked to the reader's document in terms of relevance and usability and a new situation is created in the emerging texts.

4.1.4 Drawing together the Three Investigations on the Nature of Academic Reading

All phases of the study up to this point tend to confirm that students select reading strategies to match their purposes for reading, reading expeditiously to find relevant information and reading carefully to understand and evaluate information. The diary study and follow-up interviews provided us with detailed information under the conditions of real world academic reading on the reading purposes, processes and evaluation strategies of the readers who were reading to write an assignment. Different types of careful reading associated with different cognitive processes have been established. More importantly, it has been possible to pinpoint certain characteristics of reading and combining information from multiple texts for which comprehension theories based on comprehending a single text do not sufficiently account.

As discussed in Chapter 2, propositional models in reading comprehension explain the processes that take place in the decoding of sentences and text based models explain the formation of situational models adequately and have received considerable acknowledgement in reading theory. However, the recent interest in and studies on reading multiple documents have put doubt on the validity of situational models in explaining the reading operations that take place when readers read across texts (Perfetti et al, 1999; Rouet, 2006). Recently proposed documents models emphasise that building of a situation model within and across texts do not entail the same processes as readers have to deal with possible disjunctions, discrepancies and the lack of a unifying context across the situations in separate texts and in doing so, they will resort to such strategies as source evaluation in order to build an integrated understanding of the situation. The present study has given support to the assumptions above by investigating reading strategies through three detailed studies and has presented some evidence that 'documents model' reading is identifiable as an academic reading skill as well as sentence and text level reading. It is also noted that the presence of a written assignment has largely determined the reading processes of the readers and the evaluation strategies for the inclusion of information in the assignment might well be seen as a part of the process. The Table 4.17 below summarises the suggestions from this part of the study. In brief, readers read expeditiously to find relevant information and read carefully when they deem the information as relevant. They process sentences and form immediate semantic links across sentences to form a macrostructure. They use whatever background knowledge is available to them to form a situation model through this reading. When they read further documents, they base their understanding on previously read texts and they build higher order intertextual semantic links by taking into account source characteristics. If the information is deemed trustworthy and salient, it is integrated in the previous situation model, resulting in a documents model. This documents model is unique (at least to the readers who have consequently read exactly the same documents). The reader creates a new account, a unique situation model when he or she puts this into his or her own task. The most important reading processes seen in the table are microproposition formation, forming of textual semantic links, macrostructure formation, integration of background information, intertextual semantic links, evaluation of source characteristics and usability judgements, documents model formation. These processes will be revisited when the second research question is discussed in the next section.

Table 4.17: Summary of reading for writing purposes

Reading Operation	**Purpose**	**Process**	**Outcome**
Expeditious Reading	Finding relevant information	Word, meaning matches, textual organisation clues Background knowledge on content and textual organisation	Selection of information for careful reading
Careful Reading at Text Level	Understanding the argument	Microproposition formation Textual semantic links Macrostructure formation Background knowledge of all types	Situation model formation
Careful Reading at Documents Level	Understanding and integrating multiple arguments raised on the same issue	a) Background knowledge from previous text Intertextual semantic links between texts b)Documents knowledge (Source characteristics)	Situations Model + Intertext Model = Documents model formation
Evaluation of Information	Inclusion in the task	Intertextual semantic links between the texts and the task Judgements of usability	Formation of a unique situation model

4.2 Research Question 2) Do test takers score differently on tasks operationalising three careful reading types?

The investigation of the second research question entailed a series of progressive studies that would set up the parameters for the design of the language test which would be used for the analyses of the performances of test takers in responding to items putatively assessing different types of careful reading. The details of those preliminary studies are given in Chapter 3. In this part of the chapter, the results from the verbal protocol study, which functioned as the pilot study oт the tests will be given. The findings from the verbal protocol study helped both refine the tests technically in terms of item performance and provided evidence on the congruence between the test specifications and the cognitive processes that the test takers used in attempting the questions. In that sense, the study also provided data for the third research question. After the verbal protocol study, the statistical analysis of the test data will be presented. Discussions will follow right after each section.

4.2.1 Test Takers' Verbal Protocol on Test Tasks

Before the tests were administered to a large group of test takers, they were piloted on a group of students. The participants' comments were given below as they relate to each task.

4.2.1.1 Task S

Task S was a sentence level task that putatively required matching of the titles given with the first sentences in the paragraphs. There were 25 comments identified in the data relating to the reading processes in the first task and 19 (76%) could be associated with careful reading at local level. In 7 cases among 19, the participants commented that they read ‘one or two sentences’ to match the headings. For example,

‘I am reading the headings first, then the beginnings of the paragraphs.’

‘I have read only the first or the second sentence in those paragraphs and I found the answers. These were easy.’

‘I am not going to read more because the answer is clearly in the first sentence of the paragraph.’

‘I chose the title for the paragraphs reading the first line but then I quickly read a few more lines to be sure.’

‘This is quite easy to do because it is enough to read the first lines.’

There were two comments by two participants that indicated that they skimmed the whole text before they started to answer the questions:

‘I read the text quickly to have an idea then I read the titles.’

‘I looked at all the paragraphs and I picked up the easiest paragraph and found the answer.’

Three comments by the same participant made clear that she skimmed at the paragraph level, too:

'I just look at some of the words in the paragraph.'

'I form a general idea about what it (the paragraph) is about. Yes, the 3rd question. I am trying to match heading here to see whether they mean the same thing.'

'What I do, I read the first line and look at the rest of the paragraph for key words and I go back to the heading and choose one.'

And one comment in which the participant made mention of text level, although it wasn't clear whether she meant the whole text or a paragraph:

'When I read the text, I understand the text...'

Task S was found to be quite an easy, straight forward task. The participants answered the items correctly 94% of the time. Except for two questions for one participant and one question for another, they did not mention that they had difficulty in finding the answers. In those cases, they read more to find the answer:

'Some of the headings do not match the paragraphs.'

'Why can't I find 5?'

4.2.1.2 Task T

Task T was a text level task in which the test taker was asked to choose from correct summary statements and put them into the correct order. The comments Task T received were a mixture of what the participants did and the difficulties they were having. For this reason, it would not be meaningful to report ratios; therefore the comments were grouped under three headings and reported below: comments on text level, instructions and difficulty of the task. With Task T, the participants answered the items correctly 53% of the time.

Text level: It was clear from the observations that the participants had to read the whole text and parts of the text more than once. However, only three participants commented that they started by reading the whole text. The other three said that they wanted to read the statements first and to find a matching paragraph for the first statement and thus presumably complete the task on 'matching the sentence to the paragraph' basis. However, the latter group realised that it might not be possible to do any matching before they understood the text:

'I am reading the whole text then I will read the sentences.'

'This looks like a complex task. First, I will read and understand the text.'

'I looked at the sentences. They didn't mean anything at first.... too difficult to understand when you don't know the text. Then I read everything. I went back to the paragraph where I thought a similar thing was mentioned.'

'I start by reading the statements first and I will see which paragraph they match.

'I always read the questions first, then the text. This was what I did at the beginning.'

'There are sentences to be matched with the paragraphs. So I will read the sentences first.'

During the interview, there were five comments by five different participants in which they mentioned that the whole text must be read carefully for the completion of the task:

'You have to really understand the whole texti'

'I am reading the whole text once more then.i'

'You have to read very carefully and you have to keep in mind the texti'

'I can't match the sentences to the paragraphs. I will do the other way around. I will read the text and match the paragraphs to the sentences. But first I think I need to understand the text well.'

'I read all the paragraphs in the text and I found a few answers.i'

Instructions: One important observation was that of the six participants, five could not understand the instructions clearly. These five participants made a false start by trying to make a one-to-one match between the paragraphs and the statements. The researcher asked them to read the instructions once more when they needed clarification:

'I am not sure about the first statement. I am not sure about which paragraph the first sentence is matching. Does every paragraph match a sentence?'

'I am not sure whether H correctly matches with this paragraph because the sentence matches the paragraph here (reads the relevant bit of the paragraph) *but the rest of the sentence is not in the paragraph. The*

sentence says more than what is said in the paragraph. The researcher asks the participant to read the instructions again. *Oh I see... I was reading one paragraph and trying to find the matching sentence. May be it is the influence of the first task.ὁ*

'I had problems with B, because it says it benefits both sides, but here in the paragraph it is just one side being discussed. It says local people... (shows the paragraph). The researcher asks the participant to read the instructions again. *So this sentence may be summarising two paragraphs?'* The participant moves on and reads the next paragraph to find the answer.

'This statement should be related with paragraph 2. 2 and 3 are about different approaches so there must be a sentence about each. The researcher asks the participant to read the instructions again. *Aaah, so it is not one sentence per paragraph. I was looking for a sentence for each paragraph. Then the answer should be C. Sentence C is about both approaches.ὁ*

'I need to match the statements with the paragraphs and I need to put them in the order. I am reading and reading again but I can't find anything that matches. I cannot decide which one. The researcher asks the participant to read the instructions again. *Ok ok ok, I know what you mean now. I have found one answer, but but I am not sure... I want to read the instructions again. I am going to read the first paragraph again.ὁ*

Difficulty: That the task was a difficult task was a common comment by all the participants. Besides the lack of clarity in the instructions, the participants also commented that putting the statements in order was challenging and in one case, the participant mentioned that understanding of the statements implies inferring the understanding of '*the reader's*', which the researcher interpreted as 'understanding the understanding of the text by the item writer'. This may suggest that in forming the summary of a text, a reader necessarily brings in his or her interpretation to the text and summary statements written by the item writer may reflect the interpretation of the text by the item writer, and the test taker might need to 'read the summary sentences and the text' through the item writer's eyes:

'This is a difficult task. Firstly, after the first task, you want to do the same matching here. The answers are not straight forward. You have to really understand the whole text. And then really understand the statement to see if they are matching. Because the statement is not directly from the paragraph. There is a judgement inside. It is the reader's understanding in there.è

'But now I am wondering whether to put it here or here (between the two sentences she correctly identified).*Oh, now I have found out that I have to put it in between. The first one is on 2 and 3, and the next one is about 5, so this must go in between. I don't know why but there is a task like this in another test. I take a long long time to do it, too.è*

'You have to keep in mind the passage, you have to keep in mind the options... it is a lot to do... It annoys me. It annoys me because I can't see why I can't find the answers. The text is not particularly difficult. I understand it. I understand the summary sentences, too. I have chosen the right ones but I can't match them and I can't put them in order. At the beginning, I thought I was going to find a sentence per paragraph. It put me off little bit. One sentence per paragraph would be easy. You have to read very carefully and you have to keep in mind the text. Remembering, going back to the text, checking this and checking that... You have to do so many things at the same time.'

'I try to keep in mind the sentence then I go back to the text, then I forget which letter it was. May be I should look at the text and mark the statements as the subject comes up in the text.'

'I am struggling to find the correct statement and the correct order at the same time.'

'I read the statement, I read the passage then I forget the statement and I read it again. When I decide that a statement is correct, then I have to find its place. This is difficult. '

4.2.1.3 Task I

Task I asked the participants to match the mini-texts given to the relevant paragraphs. Task I seemed to be a straightforward task and the items could be answered correctly 81% of the time. Of the six participants, five stated that they read the main text first, then the mini-texts and they could match them one by one:

‘I have to have an idea what the original article has in it. I read the text and I read the small paragraphs.’

‘Once you read the main text, it is easier to understand the mini-texts.’

‘You need to understand the text and remember the problems there in order to do this task. I remembered something was in it or not. For example this one, D, I remembered that there was a section about the financial benefits. Then when I read the text again, I matched them.’

‘There is a main text here and several paragraphs to match. I read the text first to understand it. If you know the text then you can easily see what is related to it.’

‘I had a quick look at the text and then the mini paragraphs. It was easy to see what the text may be about but not the paragraphs. Paragraphs were a bit difficult to read quickly. Then I decided to read the main text first. Then when you read the mini-texts, you remember that there was something about it in the main text.’

One participant who used search reading skills most in the study reported that she was having difficulty with this task, too.

‘The problem here is that the paragraphs (mini-texts) have all very similar key words. I cannot easily match them with the paragraphs. I cannot make

it quickly. The first sentence is about the first approach. The third one is too because I see the name of the approach in them. But the same names are in the other paragraphs, too.è

There were several comments on how the mini-texts were semantically related to the main text:

‘Oh, this is an example explaining the idea in the paragraph. They (mini-texts) are more like examples. So I could do the matching.è

‘In the text, there is an argument. And here, there is an example.è

(Reads and paraphrases the third mini-text.) ‘The balanced approach was mentioned towards the end of the text, I remember. Here, they are giving more advantages of the balanced approach.è

‘You can understand that the paragraph is about the same thing. The text mentions the harm that can be given to local culture and in the paragraph it talks about how other things, like museums, can be badly affected. It is more comments about the same thing.è

4.2.1.4 Discussion on the Verbal Protocols

Although the verbal protocol data came only from a small group of test takers, it provided helpful suggestions on the degree of congruence between the test tasks and the test specifications.

Firstly, Task S seemed to measure careful reading at local level, at most across two sentences if not one. Although some participants reported that they skimmed the text, this did not contribute to the completion of the task, nor was detrimental to it. As such, it represented a more genuine local level careful reading process, in which the reader focuses on particular information within a context, and checks his/her comprehension within the immediate context. Therefore, no concerns were raised on the design of Task S. However, sub-headings were revised to remove the ambiguity in two cases.

Secondly, Task T proved to be a truly text level task that required recursive processing of the whole text. The summary statements could not be processed on a paragraph basis; the understanding of the whole text was a prerequisite. However, the instruc ions and the sophistication of the summary statements had to be revisited to make them clearer and more explicit so that they could be related to the texts more easily. Two problem points – multiple-paragraph summarisation and ordering the information – received a lot of discussion in the focus group. It was eventually decided that if the summary statements had matched the paragraphs on a one-to-one basis, the task would allow paragraph level reading that would defeat the purpose. Since the focus group assumed that processing a text paragraph by paragraph without having to form a macroproposition of it does not equate to text level processing, the statements that summarised more than one paragraph were retained in the task. However, the utmost care was given to make the instructions as explicit as possible. The focus group rewrote the instructions so that it was clearer that the statements could summarise more than one paragraph. Ordering of the statements as the information in the text was again considered to be a necessary process in creating the summary of a text and it was retained as well. However, the layout of the task was carefully redesigned so that it would be possible for test takers to see the text, the statements and the box where they would record their answers.

Thirdly, Task I appeared to have been operationalised successfully as specified in the task descriptions and the test specifications. The participants mentioned the necessity of reading carefully both the main text and the mini-texts for the successful completion of the task. It was mentioned that for understanding how the mini-texts were related to the main text, the main text had to be read and understood. One participant who tried to do the reading expeditiously commented that she failed to match the paragraphs to the texts by reading quickly and matching words. This was important as the test takers were expected to build a situation model of the first text so that they could process the new information based on the background information that the first text provided to evaluate the relation of the new information to what they already learned from the first text. The comments above supported that that was what the participant test takers had done. The participants also commented on meaning (semantic) relations between the main text and the mini-texts;

e.g.: the mini-text provided more information on the same subject by discussing another aspect of the subject, or exemplified a case raised in the main text. In doing so, the participants exhibited that they understood the relevance of the incoming information to what they had already read and constructed intertextual connections taking into consideration how the incoming information complemented the information in the text. As it stood, Task I appeared to activate the expected cognitive processes of building semantic links between the texts (intertextual proposition formation) as they were described in the test specifications and as they are discussed in the literature of documents model formation (i.e.: Perfetti et al., 1999). In one case, explicit word match between the text paragraph and the mini-text was removed.

In sum, the verbal protocol data revealed that Task S was a local level careful reading task which could be easily completed successfully by reading the first sentence or sentences of the paragraphs. Task T, on the other hand, proved to be quite a challenging task that required recursive reading and simultaneous processing of text information, text information organisation and the information in the summary sentences. The protocol data implied several improvements to the task, some of which were implemented before the test administration. The improvements were expected to decrease the difficulty with which Task T could be completed. On the other hand, Task I appeared to have been designed successfully. Considering that the participants in the study were advanced level students, 81% success rate was deemed appropriate for Task I. After the final revisions, the test was administered to a large group of students for the statistical analysis of test taker performance on the tasks (see Section 4.2.2 below).

4.2.2 Statistical Analyses of Test Results

The second research question investigated whether there were significant differences in the scores students achieved in tests designed to operationalise different types of careful reading. It is hypothesised that each task type would make differing demands on students, which would be reflected in different levels of performance and different patterns of response. It is expected that testlets each with their own texts and items operationalised at sentence (Task S), text (Task T) and multiple text (intertextual) levels (Task I) would reflect a difficulty continuum (see Section 2.4). In order to analyse test performance, descriptive statistics were obtained, reliability estimates, item analyses (corrected item total correlation, CITC; alpha if item deleted, AIID; and item discrimination patterns IDPs) and mean comparisons (ANOVA) were performed and these are reported below. Embedded in the investigation, there was also the issue of whether there was congruence between the dimensions of the reading construct as reflected in the test specifications and in the test (see Appendix 3.5 for test specifications). The analysis of the internal structure of tests (careful reading at sentence, text and intertextual levels) is reported below through Principal Component Analysis (PCA) to investigate whether the items putatively measuring different careful reading components load on different factors.

4.2.2.1 Descriptive Statistics of Test Versions

Table 4.18 below shows the mean scores obtained on three test versions on the basis of the following tasks; Task S, careful reading at sentence level (matching the headings to paragraphs); Task T, careful reading at text level (selecting correct summary statements and putting them into order); Task I, careful reading at intertextual level (matching content-wise related mini-texts to the paragraphs in the text). The test takers performed differentially in these three tasks, scoring highest in Task S (62.71%) and lowest in Task T (37.97%). The normality tests were run to check whether the normality assumption required in further analyses was violated or not. The score distribution in Task Sand Task T is near normal according to skewness and kurtosis values (skewness= ± 1, kurtosis= between -1 and 2). In Task I, the kurtosis value is marginally below -1 showing that scores are more even giving a slightly playkurtic (flat) distribution (see Figures 4.1,4.2, 4.3). However, Kolmogorov-Smirnov tests fail to support normality ($p< .0005$) (Table 4.19). Pallant (2007) suggests that with larger samples this is quite normal, in which case inspection of Normal Q-Q plots would be helpful, and in this case the plots supported normality (see Appendix 4.1 for Normal Q-Q plots).

Table4.18: Descriptive statistics of scores in three versions (N=1032)

Task	Mean (%)	Std. Dev. (%)	Skewness	Kurtosis
Task S	62.71	29.84	-.292	-.958
Task T	37.97	30.31	.608	-.489
Task I	49.32	33.21	.147	-1.170

Table 4.19: Kolmogorov-Smirnov tests

	Task S (%)	Task T (%)	Task I (%)
Kolmogrov-Smirnov Z	4.698	6.489	4.982
Sig.	<.0005	<.0005	<.0005

Figure 4.1: Task S score distribution

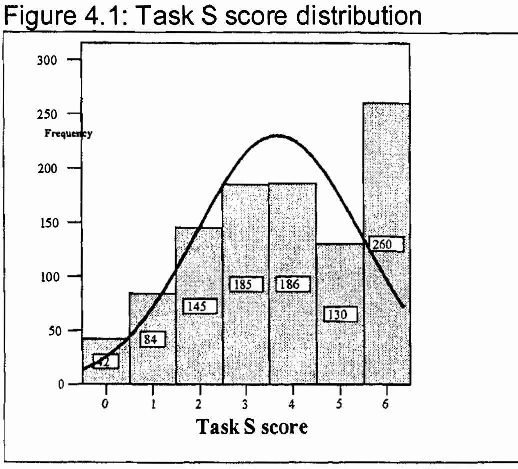


Figure 4.2: Task T score distribution

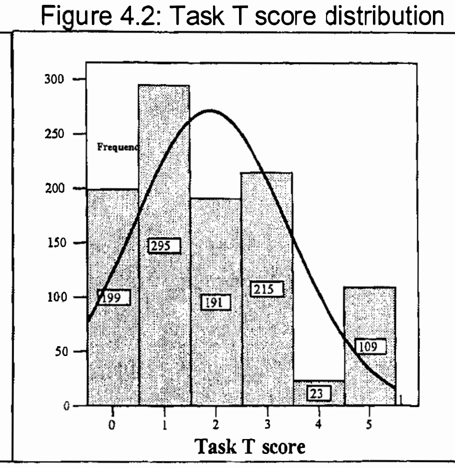
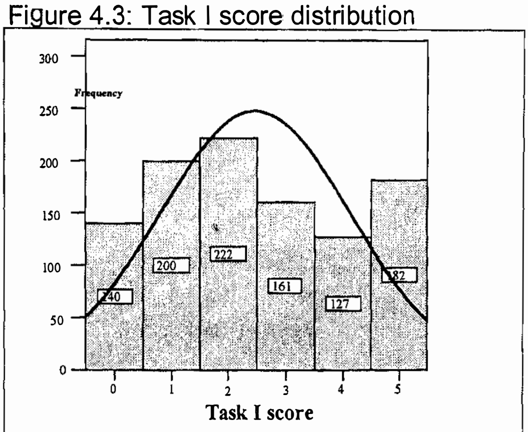


Figure 4.3: Task I score distribution



When each version (Version 1, 2 and 3) was analysed separately, the means and score distribution of the tasks in each version are revealed as in Tables 4.20. In all the versions, Task T has the lowest mean, and except for Version 2, Task S has the highest mean among the three tasks. In all the test versions, skewness values are within the range of -1 and +1s and kurtosis values are only slightly above -1 in the third tasks. Although Kolmogorov-Smirnov tests did not support normality (Table 4.21), Normal Q-Q plots revealed a near normal distribution (see Appendix 4.2).

Table4.20: Descriptive statistics of scores

Version	Task	Mean (%)	Std. Dev. (%)	Skewness	Kurtosis
Version 1 (N: 340)	**V1TS**	72.40	26.96	-.557	-.855
	V1TT	42.82	30.79	.313	-.791
	V1TI	47.35	32.86	.244	-1a1a
Version 2 (N: 347)	**V2TS**	43.76	26.72	.228	-.603
	V2TT	39.25	32.83	.624	-.665
	V2TI	55.33	32.69	-.132	-1.15
Version 3 (N: 345)	**V3TS**	72.22	26.24	-.634	-.411
	V3TT	31.88	25.95	.819	.221
	V3TI	45.22	33.32	.343	-1a05

Table 4.21: Kolmogrov-Smirnov tests on versions and tasks

Version/Task	Kolmogorov-Smirnov Statistics	df	Sig.
V1TS	.212	340	.000
V1TT	.185	340	.000
V1TI	.162	340	.000
V2TS	.130	347	.000
V2TT	.185	347	.000
V2TI	.149	347	.000
V3TS	.197	345	.000
V3TT	.233	345	.000
V3TI	.171	345	.000

4.2.2.2 Internal Reliability Estimates and Item Statistics in Test Versions

In this section, internal reliability estimates and item statistics are presented by tasks across test versions. In Table 4.22, Cronbach's alpha estimates for Task S across three versions (.703, .617, .671 respectively) are moderately high for a testlet of six questions. Table 4.23 gives item (item facility and standard deviation) and item-total statistics (corrected item total correlation and alpha if item deleted) of Task S across three versions. The item facility (mean) values range from 49-89% in Version 1, 17-69% in Version 2 and 53-90% in Version 3. The item-total statistics revealed two problematic items, V2TS-4 and V2TS-6, which presumably contributed to the low mean of the Task S of Version 2. When the items were reanalysed, some technical problems in the operationalisation of the items that did not surface in the verbal protocol phase were detected. With V2TS-4, the correct heading was not explicitly related to the first sentence of the paragraph, and there was a word overlap with another heading and the first line of this paragraph. This resulted in many cases where the latter heading was chosen as the answer.

Similarly, the first sentence of the paragraph with which V2TS-6 was to be matched had a word overlap with a distractor, making the distractor an alternatively correct answer for that paragraph. Therefore, V2TS-4 and V2TS-6 were discarded in the further analyses. Another item, V1TS-2, had negative effect on the reliability of the task (AIID: .720). However, as no technical problems could be identified with the item, it was retained in the task as a relatively easy item. Item discrimination (IDP) graph of V2TS-6 confirmed an erratic score distribution among lower and higher performing groups, and there was a slight dip from the 4th band to the highest band in the IDP graph of V2TS-7.Otherwise, IDPs of Task S items did not reveal any problematic distribution of scores to lower and higher groups in terms of their score performance in the tests (see Appendix 4.3 for IDPs).

Table 4.22: Reliability statistics of Task S across three versions

Task S - Version	Cronbach's Alpha	N of Items
1	.703	6
2	.617	6
3	.671	6

Table 4.23: Item, item-total statistics of Task S across versions

Version/Task-Item	Mean	Std. Dev.	CITC*	AIID**
V1TS-2	.85	.360	.217	.720
V1TS-3	.62	.485	.588	.606
V1TS-4	.49	.501	.540	.625
V1TS-5	.69	.464	.475	.649
V1TS-6	.81	.391	.452	.658
V1TS-7	.89	.316	.329	.693
V2TS-2	.69	.465	.447	.531
V2TS-3	.68	.467	.443	.533
V2TS-4	.17	.374	.469	.536
V2TS-5	.34	.473	.275	.603
V2TS-6	.28	.448	.151	.646
V2TS-7	.48	.500	.352	.572
V3TS-2	.72	.452	.407	.627
V3TS-3	.79	.407	.397	.631
V3TS-4	.84	.367	.393	.633
V3TS-5	.56	.498	.381	.640
V3TS-6	.90	.302	.284	.664
V3TS-7	.53	.500	.553	.567

* Corrected item-total correlation
** Alpha if item deleted

Cronbach alpha values for Task T were .637, .757, .566 in the three versions respectively (Table 4.24). These coefficients are moderately high for a five-item task, the coefficient for Version 3 being the lowest. Table 4.25 shows facility values and item-total statistics in Task T across versions. Item facility values (mean) range between 28-64% for Version 1, 21-68% for Version 2 and 8-62% for Version 3. Task T, with the lowest means across versions did not exhibit any item related statistical problems except for extremely low item facility of V3TT-6 (8%). When the items were revisited, it was seen that in Version 1, one of the two distractors was too strong (correct answer for V1TT-2 was replaced by the distractor almost 60% of the time). The distractor statement included a detail that was not stated in the text, however, it stood as an attractive and generally plausible statement given the context of the subject matter in the text; the detail in the statement that was not supported by the text was too often disregarded. In Version 2, two correct summary statements (V2TT-4, V2TT-6) were incorrectly used instead of each other; they were matched with alternated paragraphs and thus put in a wrong order, too. One summary statement suggested that 'both *methods* should be used', and the other, 'neither of the methods is better than the other'. Apparently, these statements were too close in meaning and had to be detailed for a better match with the text; however the meaning difference was explicit. In Version 3, which yielded the lowest mean (31.88%) among other Task Ts, the distractors distracted the test takers too strongly. Although neither of the statements were correct summaries of the main ideas in the text, they could be matched with minor details stated in the paragraphs and were chosen as correct answers too frequently in the last two items in the task. However, in terms of item discrimination across low and high performing groups, IDPs did not show any unfavourable item performance except for V1aT-3 which did not successfully discriminate the test takers only at the third and fourth bands (see Appendix 4.4).

Table 4.24: Reliability statistics of Task T across three versions

Task T - Version	Cronbach's Alpha	N of Items
1	.637	5
2	.757	5
3	.566	5

Table 4.25: Item, item-total statistics of Task T across versions

Version/Task-Item	Mean	Std. Dev.	CITC	AIID
V1TT-2	.37	.483	.306	.625
V1TT-3	.28	.449	.386	.587
V1TT-4	.64	.481	.318	.619
V1TT-5	.46	.499	.515	.518
V1TT-6	.40	.491	.434	.562
V2TT-2	.68	.466	.413	.753
V2TT-3	.36	.480	.570	.697
V2TT-4	.28	.448	.578	.695
V2TT-5	.43	.496	.463	.738
V2TT-6	.21	.410	.622	.684
V3TT-2	.62	.486	.251	.561
V3TT-3	.41	.492	.363	.488
V3TT-4	.24	.428	.467	.425
V3TT-5	.25	.433	.301	.524
V3TT-6	.08	.269	.290	.539

* Corrected item-total correlation
** Alpha if item deleted

Table 4.26 below presents reliability coefficients of Task I across three versions; .676, .692, .706 respectively, moderately high values for a five-item task. In general, Task I is more uniform across the three versions both in terms of facility values which range between 43-55% for Version 1, 41-69% for Version 2 and 38-59% for Version 3 and it is also quite uniform in terms of item-total statistics (see Table 4.27). No items appeared problematic either statistically or qualitatively and IDPs supported this except for V2TI-2, which did not discriminate between the test takers in the middle and highest group (see Appendix 4.5).

Table 4.26: Reliability statistics of Task I across three versions

Task I - Version	Cronbach's Alpha	N of Items
1	.676	5
2	.692	5
3	.706	5

Table 4.27: Item, item-total statistics of Task I across versions

Version/Task-Item	Mean	Std. Dev.	CITC	AIID
V1TI-2	.47	.500	.499	.593
V1TI-3	.55	.499	.406	.635
V1TI-4	.49	.501	.423	.627
V1TI-5	.43	.496	.416	.631
V1TI-6	.43	.496	.401	.637
V2TI-2	.69	.461	.482	.628
V2TI-3	.59	.492	.416	.657
V2TI-4	.48	.500	.477	.629
V2TI-5	.41	.493	.384	.669
V2TI-6	.59	.493	.482	.627
V3TI-2	.59	.492	.425	.672
V3TI-3	.37	.485	.582	.607
V3TI-4	.49	.501	.513	.635
V3TI-5	.38	.485	.352	.701
V3TI-6	.43	.495	.446	.663

* Corrected item-total correlation
** Alpha if item deleted

All in all, the item analyses above pointed out two items in Version 2 Task S with detectable technical problems that violated the test specifications and these items were eliminated in further analyses. Besides, further analysis of Task T responses as well suggested ways with which performance in Task T could be improved. However, no items were eliminated as this would negatively impact the statistical performance of the tests except for increasing the means for Task T. After the initial analyses above, the data were analysed to see whether there were any statistically significant differences between the task mean scores, and whether there is any text (in this case version) or a combined task-text effect on the scores. The next section will present ANOVA analysis for the investigation of above stated issues.

4.2.2.3 Mean comparisons

One-way repeated measures ANOVA

The first step of ANOVA analysis was the comparison of test scores on the three test tasks (Task S, Task T and Task I) as dependent variables and test scores as independent variable using one-way repeated measures. As each test taker took Task S, Task T and Task I consecutively, one-way repeated measures ANOVA was deemed appropriate in the analysis of general performance in tasks. Although Kolmogorov-Smirnov test had shown that none of the test score distributions was normal (see Section 4.2.2.1), it is assumed that the procedure could be used in analysing the data as one-way ANOVA's F test is robust for validity against the violation of this assumption (Pallant, 2007). The sphericity assumption was satisfied ($p= .196$). The defective items V2TS-4 and V2TS-6 were excluded from the analysis (see Section 4.2.2.2). The means and standard deviations are given in Table 4.28 and ANOVA results in Table 4.29. There was a significant effect for Task, (Wilk's Lambda = .605, $F(2, 1030) = 251.38$ $p<.0005$) and estimated marginal means graph (Figure 4.4) represented the differences in among the means of the three tasks.

Table 4.28: Descriptive statistics excluding items V2TS-4 and V2TS-6

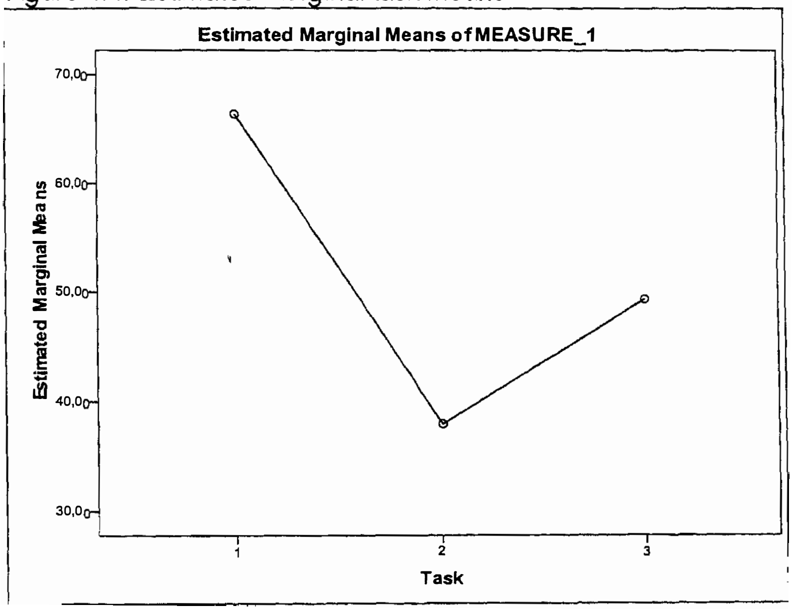
Task	N	Mean	Std. Dev.
Task S	1032	66. 35	29.54
Task T	1032	37.97	30.31
Task I	1032	49.32	33.21

Table 4.29: ANOVA one-way repeated measures

N	Wilk's Lambda	F	df	*p* (Sig.)	Partial Eta Squared (Effect Size)
1032	.605	336.39	1030	.000	.395

.01: small effect, .06: moderate effect, .14: large effect

Figure 4.4: Estimated marginal task means



One-way repeated measures ANOVA: Post hoc tests

After obtaining a significant difference in ANOVA analysis, in order to identify whether there was a statistically significant difference between each set of scores, post hoc tests were conducted using pairwise comparisons. Since three t-tests (Task S & T, Task T & I and Task S & I) were involved, a more stringent *p* value was chosen for the analysis; the *p*-values were taken as statistically significant at .0167 level, according to Bonferoni adjustment (.05/3). The results indicated significant differences among the scores from the tasks across versions, with large effect size (see Table 4.30).

Table 4.30: Pairwise comparisons (Sig. at .0167 level)

Task	Mean difference	Sig.	Partial Eta Squared (effect size)
Task S X Task T	±28.37	.000	.395
Task S X Task I	±17.01	.000	.395
Task T X Task I	±1[illegible]36	.000	.395

Adjustment for multiple comparisons: Bonferroni
.01esmall effect, .06: moderate effect, .14: large effect

A mixed between-within subjects ANOVA

In order to determine the source of the mean difference and further analyse any differential contribution of the tasks and the versions, a mixed between-within subjects ANOVA was conducted. Therefore, the first set of independent variables was the tasks; Task S, Task T, Task I as within-subjects factor since all the test takers were given all the three tasks. The second set of independent variables was versions; Version 1, Version 2, Version 3 as between-subjects factor since the test takers were grouped into three according to the version of the test they took. The test scores in percentages were dependent variable. The score distribution was not normal either in tasks or versions. However, because ANOVA's F test is considered as robust for validity against this assumption (Pallant, 2007), the procedure was used in the analysis.

The homogeneity of variance assumption (Levene's test of equality of error variances) in the within-subjects data was violated in two cases (Task S: F(2, 1029) p=.000 and Task T: F(2, 1029) p= .000) and was satisfied in the third (Task I: F(2, 1029) p= .999). ANOVA is reasonably robust to violations of the assumption provided the size of the groups is reasonably similar and this was the case in the present data (N= 340, 347, 345; Versions 1-3 respectively). Homogeneity of inter-correlations was violated (Box' M statistics: p=.000). Sphericity assumption for within-subject analysis was satisfied (Mauchly's test of sphericity: p= .741). Due to the violation of homogeneity of variance assumption, a more conservative *p* value for determining significance was set; .01e(Pallant, 2007). The means and standard deviations for each task in each version are given in Table 4.31ȩANOVA results are given in Table 4.32 and Figure 4.5 represents the estimated marginal means across tasks and versions.

The mixed between-within subjects ANOVA indicated a statistically significant main effect for *task* (Wilk's Lambda: F[2,1028]=363.802, p<.0005), and a statistically significant interaction effect for *task by version* (Wilk's Lambda: F[2,1029]=43.205, p<.0005) (see Table 4.32)e The main effect for *version* was not found to be significant at the set level (Wilk's Lambda: F[2, 1029]= 4.161, p= .016). The effect size for task was very large (partial eta squared: 261) and the effect size was moderate for task by version interàction (.077). Although the analysis indicated an interaction effect making the interpretation of the main effect less straight forward, the plot of estimated marginal means (see Figure 4.5) did reveal a similar pattern across tasks suggesting that task factor might be largely accountable for the differences in the score distribution. The strength of task effect size (i.e. large effect size) was also supportive of this finding.

However, there were differences among the mean scores of the same tasks across the versions; the most prominent differences having appeared between the tasks of Version 2 and the others. Especially Version 2 Task S (V2TS), which had two defective items excluded, differed substantially from other Task Ss, i.e. V1TS and V3TS. In Figure 4.5, the plot that represents the task means in Version 2 crosses Version 1 and Version 3 mean plots, whereas Version 1 and Version 3 run in a reasonably parallel form to each other. This is also indicative of the fact that Version 2 with differences in its structure might be adding to task-version interaction effect.

Table 4.31: Descriptive statistics by task excluding items V2TS-4 and V2TS-6

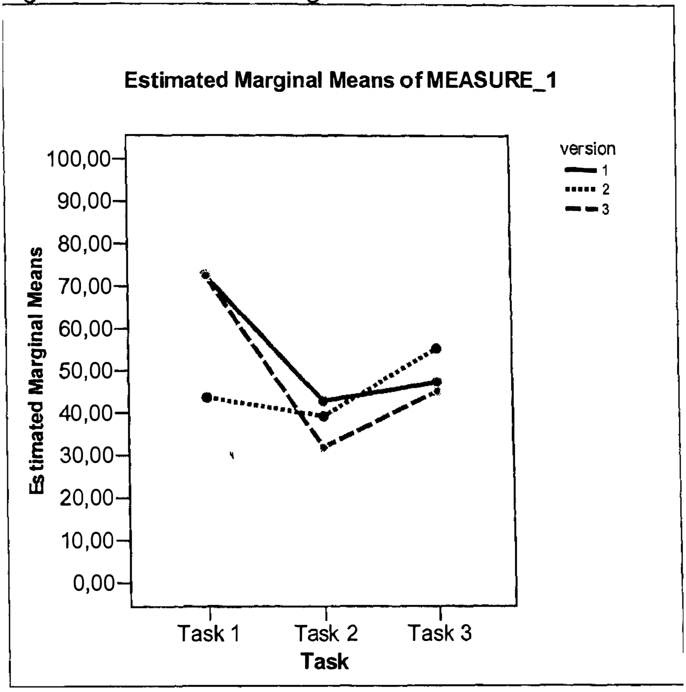
Version	**Task**	**Meant(%)**	**Std. Dev.t(%)**
Version 1 (N: 340)	**V1TS**	72.40	26.96
	V1TT	42.82	30.79
	V1fTI	47.35	32.86
Version 2 (N: 347)	**V2TS**	54.54	31.51
	V2TT	39.25	32.83
	V2TI	55.33	32.69
Version 3 (N: 345)	**V3TS**	72.22	26.24
	V3TT	31t88	25.95
	V3TI	45.22	33.32

Table 4.32: Mixed between-within subjects ANOVA (sig. at .01 level)

Factor	**F**	**df**	***p* (sig.)**	**Partial Eta Squared (effect size)**
Task	363.802	1029	.000	.261
Version	4.161	1029	.016	.008
Task*Version	43.205	1029	.000	.077

.01: small effect, .06: moderate effect, .14: large effect

Figure 4.5: Estimated marginal means across tasks and versions



A mixed between-within subjects ANOVA: post hoc tests

Following the finding of a substantial effect for task, and a moderate task-version interaction effect on test scores, several post hoc tests were carried out to help the interpretation of the main effect on test scores.

Although the ANOVA did not reveal *version* as a statistically significant factor, because there was a task-version interaction effect, the versions alone were compared using pairwise post hoc tests of mixed between-within subjects ANOVA (Table 4.33). Since three comparisons were done on version, Bonferroni correction was applied to the significance level to set *p* at .017 (.05/3) value. No significant differences between the versions were found at the set level and the effect sizes were negligibly small.

Next, planned comparisons for task-version interaction effect were conducted using paired samples t-tests for tasks and independent samples t-tests for versions (Table 4.34 and 4.35 respectively). As nine tests were used in task-version interaction analysis, the *p* value was set at the .0055 level (.05/9) (Pallant, 2007). Task-version interaction was analysed firstly by comparing tasks in each version (Table 4.34), secondly by comparing versions across tasks (Table 4.35). In comparisons on tasks where significant differences were expected, there were two cases in which significant differences were not detected. The difference between Task T and Task I in Version 1 was not significant at the set level, and the difference between Task S and Task I in Version 2 was not significant either, the former (V2TS) being the problematic test task. Otherwise, the means of Task S, Task T and Task I in all the versions were significantly different from each other with large effect sizes. In comparing versions, where we would expect the same tasks (e.g. V1TS vs. V2TS vs. V3TS) yield non-significant mean differences across versions, there found to be six statistically significant mean differences out of nine cases (Table 4.35). However, among these six cases only three had moderate to large effect sizes (Version 1&2 of Task S, Version 2&3 of Task S and Version 1&3 of Task T), two of which involve the problematic Version 2 Task S (V2TS).

Table 4.33: Between-subjects factor pairwise comparison (Sig. at .0167 level)

Version	Mean difference	Sig.	Partial Eta Squared (effect size)
Version 1 X Version 2	±4.49	.035	.008
Version 1 X Version 3	±4.42	.040	.008
Version 2 X Version3	±.068	1.00	.008

.01esmall effect, .06: moderate effect, .14: large effect

Table 4.34: Post hoc, paired samples *t*-tests on tasks [sig. at .0055 level]

	Task S & T		Eta Sq. effect size	Task T & I		Eta Sq. effect size	Task S & I		Eta Sq. effect size
	t	*p*		t	*p*		t	*p*	
Version1	16.53	.000	.45	-2.25	.025	.01	13.74	.000	.36
Version2	7.73	.000	.15	-9.11	.000	.20	-.43	.670	.00
Version3	24.29	.000	.64	-7.48	.000	.14	14.70	.000	.39

01: small effect, .06: moderate effect, .14: large effect

Table 4.35: Post hoc, independent samples *t*-tests on versions [sig. at .0055 level]

	Ver 1&2		Eta Sq. effect size	Ver 2&3		Eta Sq. effect size	Ver 1&3		Eta Sq. effect size
	t	*p*		t	*p*		t	*p*	
Task S	7.99	.000*	.08	-8.02	.000*	.08	.088	.930	.00
Task T	1.47	.142	.00	3.28	.001*	.02	5.03	.000*	.04
Task I	-3.19	.001	.01	4.03	.000	.02	.845	.399	.00

*Equal variance not assumed
.01esmall effect, .06: moderate effect, .14: large effect

In sum, the ANOVA above showed that the tasks differed from each other significantly in terms of their mean scores - task difference has been confirmed as a strong factor in the data, and the versions, i.e. counter balanced matching of texts with tasks did not generally create any significant difference. Task-version interaction effect was generally accounted for by the dissimilar performance of Version 2. Therefore, it can be suggested that the tasks generally functioned differentially, Task S receiving higher scores than Task T in all cases, the score on Task S being also significantly higher than that on Task I in two out of the three versions (except for in Version 2, where facility values for Task S were relatively low and those for Task I relatively high). Therefore, the items which target sentence level comprehension (Task S) seem to receive higher scores than the items which target understanding the whole text (Task T) in all cases, and those which target linking of two texts (Task I) obtained higher means than the items which target understanding the whole text (Task T) in all versions.

4.2.2.4 Analysis of the Internal Structure of the Tests

The next step in the study was the analysis of the internal structures of the tests to investigate whether the items putatively measuring different careful reading components load on different factors.

An initial analysis of the overall and by version correlations between the tasks showed that although all the tasks are significantly correlated, none of the correlations (ranging from .301 to .497) are particularly high (see Table 4.36). To ensure accuracy in analysis, two methods, *Principal Component Analysis* (PCA) and *Factor Analysis* (Principal Axis Factoring, PAF) were computed using orthogonal (varimax), and oblique (direct oblimin and promax) rotation techniques for the investigation of the internal structure of the tests. While direct oblimin rotation technique did not reveal easily interpretable factor structures (e.g. some items not loading on any factors, some others loading on multiple factors), PCA with varimax and PAF with promax rotation yielded more interpretable and very similar results. For the reader's convenience, the results from PCA with varimax rotation are reported here and the ones from PAF with promax rotation are given in Appendix 4.6.

Table 4.36: Task correlations

Pearson Correlations	Task S&T		Task T&I		Task S&I	
	r	*p*	r	*p*	r	*p*
Overall	.309*	.000	.406*	.000	.333*	.000
Version 1	.352*	.000	.324*	.000	.382*	.000
Version 2	.344*	.000	.497*	.000	.421*	.000
Version 3	.301*	.000	.397*	.000	.362*	.000

*significant at .01 level (2-tailed)

Version 1

The 16 items from Version 1 were submitted to PCA without constraining the number of factors to be extracted. KMO measure of sampling adequacy was .734 and Bartlett's test of sphericity was significant at .000 level, both of which were above the minimum required levels. No communalities below .300 were observed. Five components with eigenvalues higher than 1.00 were extracted and these accounted for 59.05% of variance in the data (see Appendix 4.7 for the details). The rotated component matrix is given in Table 4.37 in which loadings of the items on the components that are above .300 are marked in bold and the amount of variance accounted by each component is given in parentheses.

Table 4.37: Version 1 Rotated component matrix*

Item	Component				
	1 (24.29%)	2 (10.51%)	3 (9.22%)	4 (8.54%)	5 (6.49%)
V1TS-2	.056	.056	-.060	.106	**.623**
V1TS-3	.164	.194	**.360**	-.010	**.699**
V1TS-4	.225	.093	**.432**	-.010	**.585**
V1TS-5	.183	.135	**.601**	-.064	.246
V1TS-6	.003	.136	**.707**	.040	.193
V1TS-7	.086	.022	**.741**	.103	-.102
V1TT-2	.063	.010	.044	**.915**	.005
V1TT-3	.011	.081	.034	**.900**	.117
V1TT-4	.020	**.569**	.240	.020	.272
V1TT-5	.197	**.854**	.103	.055	.092
V1TT-6	.114	**.853**	.012	.033	.013
V1TI-2	**.743**	.046	-.033	-.034	.251
V1TI-3	**.545**	.297	.261	.078	-.194
V1TI-4	**.683**	.065	-.038	-.046	.165
V1TI-5	**.633**	.011	.141	.005	.139
V1TI-6	**.562**	.169	.188	.203	-.112

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
*Rotation converged in 6 iterations.

Table 4.37 shows that Component 1 clearly accounts for Task I. First two items of Task T (V1TT-2 and V1TT-3) load on Component 4 and the rest of the items of Task T load on Component 2. The items from Task S load on Component 3 and Component 5, with two items (V1TS-3 and V1TS-4) loading on both of the components. Component 4 accounts for only two items. Using Parallel Analysis, it was found that eigenvalues of only four factors exceeded the corresponding criterion values for a randomly generated data matrix of the same size (see Table 4.38). The scree plot revealed a break after the fourth component as well (see Appendix 4.7). Therefore, the PCA was repeated by constraining the number of factors to four. Rotated 4-component matrix is given in Table 4.39 in which loadings of the items on the components that are above .300 are marked in bold and the amount of variance accounted by each component is given in parentheses.

Table 4.38: Comparison of eigenvalues from PCA and criterion values from parallel analysis

Component Number	Actual Eigen Value from PCA	Criterion Value from Parallel Analysis	Decision
1	3.887	1.3933	accept
2	1.681	1.3005	accept
3	1.476	1.2425	accept
4	1.367	1.1867	accept
5	1.038	1.1397	reject

Table 4.39: Version 1 Rotated component matrix*
4 components retained

Item	Component			
	1 (24.29%)	2 (10.51%)	3 (9.22%)	4 (8.54%)
V1TS-2	**.332**	.105	.014	.072
V1TS-3	**.715**	.189	.156	-.034
V1TS-4	**.702**	.237	.065	-.025
V1TS-5	**.632**	.159	.132a	-.056
V1TS-6	**.682**	-.031	.139	.054
V1TS-7	**.530**	.029	.046	.133
V1TT-2	.043	.063	.009	**.914**
V1TT-3	.104	.021	.073	**.893**
V1TT-4	**.368**	.027	**.556**	.013
V1TT-5	.158	.202	**.848**	.051
V1TT-6	.037	.149	**.851**	.030
V1TI-2	.134	**.759**	.028	-.047
V1TI-3	.103	**.514**	.315	.097
V1TI-4	.079	**.694**	.053	-.054
V1TI-5	.203	**.630**	.005	.004
V1TI-6	.093	**.540**	.180	.215

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
*Rotation converged in 5 iterations.

PCA with four factors accounted for 52.56% of the variance in the data and yielded a clearer distribution of the items to the components as it was possible to explain Task S and Task I with single components. However, Task T remained ambiguous; the first two items loading strongly on a separate component which did not account for any other items in the test.

To confirm the relation between the tasks and the components, a correlation analysis was run between the factor scores saved as variables and the total score of each task. Table 4.40 below gives the correlations between the factor scores and the total task scores and this confirms that there is a very high correlation between Task S and Component 1 and Task I and Component 2. On the other hand, Task T primarily correlates with Component 3 and secondarily with Component 4.

Table 4.40: Pearson task/component correlations on Version 1

Component	Task S	Task T	Task I
1	**.956***	.222*	.185***
2	.197*	.140***	**.952***
3	.151**	**.744***	.176***
4	.021	**.579***	.065

*Correlation is significant at the .000 level (2-tailed)
**Correlation is significant at the .005 level(2-tailed)
***Correlation is significant at the .01 level (2-tailed)

Version 2

The purged form of Version 1 with 14 items, excluding two defective items (V2TS-4 and V2TS-6) was submitted to PCA without constraining the number of factors to be extracted. KMO measure of sampling adequacy was .833 and Bartlett's test of sphericity was significant at .000 level, both of which were above the minimum required levels. No communalities below .300 were observed except for V2TS-5 (.224). Three components with eigenvalues higher than 1.00 were extracted and these accounted for 48.83% of variance in the data (see Appendix 4.8 for the details). Rotated component matrix is given in Table 4.41 in which loadings of the items on the components that are above .300 are marked in bold and the amount of variance accounted by each component is given in parenthesis.

Table 4.41: Version 2 Rotated component matrix*

ITEM	COMPONENT		
	1 29.78%	2 10.48%	3 8.58%
V2TS-2	.049	.047	**.839**
V2TS-3	.122	.049	**.826**
V2TS-5	**.389**	.139	.231
V2TS-7	**.393**	.215	**.386**
V2TT-2	.037	**.657**	-.036
V2TT-3	.265	**.715**	.043
V2TT-4	.195	**.730**	.057
V2TT-5	.299	**.561**	.088
V2TT-6	.128	**.746**	.228
V2TI-2	**.642**	.220	.093
V2TI-3	**.593**	.143	.042
V2TI-4	**.621**	.225	.138
V2TI-5	**.630**	.098	-.085
V2TI-6	**.682**	.105	.154

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
*Rotation converged in 5 iterations.

In Table 4.41a it is seen that Task T and Task I are accounted for by Component 2 and 1 respectively. However, Task S exhibits a rather complex structure with two items loading on Component 3, one item on Component 1 and one item loading on both Component 1 and 3. Task S in this version had only four items as opposed to other Task Ss as two defective items were excluded from the analyses. Although a three-component model would be more meaningful for a three-task test, as the Scree plot indicated two components and parallel analysis also confirmed only two components with eigenvalues exceeding the corresponding criterion values (see Table 4.42). , a two-component PCA was run as well (see Appendix 4.8 for details)

Table 4.42: Comparison of eigenvalues from PCA and criterion values from parallel analysis

Component	Actual Eigenvalue From PCA	Criterion Value From Parallel Analysis	Decision
1	4.169	1.3851	accept
2	1.467	1.2962	accept
3	1.201	1.2316	reject

The PCA analysis with two components retained yielded a less interpretable distribution of the items to the components and naturally less variance (40.26%) accounted for in the data (see Table 4.43). It especially made the interpretation of Task I more difficult, which was otherwise very straightforward. Therefore, for further considerations of Version 2, the three-factor structure was taken into consideration.

Table 4.43: Version 2 Rotated component matrix*
2 components retained

Item	Component	
	1 (29.78%	2 (10.48%)
V2TS-2	-.061	**.760**
V2TS-3	-.020	**.780**
V2TS-5	.284	**.363**
V2TS-7	**.327**	**.493**
V2TT-2	**.577**	-.093
V2TT-3	**.733**	.072
V2TT-4	**.707**	.052
V2TT-5	**.614**	.146
V2TT-6	**.686**	.193
V2TI-2	**.505**	**.344**
V2TI-3	**.421**	.286
V2TI-4	**.491**	**.374**
V2TI-5	**.422**	.196
V2TI-6	**.419**	**.430**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
*Rotation converged in 3 iterations.

To confirm the relation between the tasks and the components, a correlation analysis was run between the three factor scores saved as variables and the total score of each task as it was done with Version 1.

Table 4.44 below gives the correlations between the factor scores and the total task scores, which reveals that Task T and Task I are well correlated with Component 2 and 3 respectively and Task S is primarily correlated with Component 3 and secondarily with Component 1.

Table 4.44: Pearson task/component correlations on Version 2

Component	Task S	Task T	Task I
1	**.366***	.276*	**.946***
2	.173*	**.951***	.236*
3	**.855***	.102*	.102*

*Correlation is significant at the .01 level (2-tailed)

Version 3

The 16 items from Version 3 were submitted to PCA without constraining the number of factors to be extracted. KMO measure of sampling adequacy was .802 and Bartlett's test of sphericity was significant at .000 level, both of which were above the minimum required levels. No communalities below .300 were observed. Four components with eigenvalues higher than 1.00 were extracted and these accounted for 48.3% of variance in the data (see Appendix 4.9). The rotated component matrix is given in Table 4.45 in which loadings of the items on the components that are above .300 are marked in bold and the amount of variance accounted for by each component is given in parentheses.

Table 4.45: Version 3 Rotated component matrix*

Item	Component			
	1 (25.51%)	2 (10%)	3 (8.05%)	4 (6.73%)
V3TS-2	**.560**	.099	.079	.204
V3TS-3	**.557**	.154	-.064	.156
V3TS-4	**.654**	.022	.104	-.184
V3TS-5	**.578**	.114	.200	-.091
V3TS-6	**.522**	-.016	-.122	.111
V3TS-7	**.703**	.218	.072	.082
V3TT-2	.024	.102	.058	**.785**
V3TT-3	.256	.257	.273	**.530**
V3TT-4	.113	.183	**.628**	**.328**
V3TT-5	.075	-.019	**.790**	-.008
V3TT-6	-.034	.164	**.633**	.022
V3TI-2	-.039	**.683**	.030	.140
V3TI-3	.213	**.718**	.132	.084
V3TI-4	.154	**.688**	.087	.176
V3TI-5	.052	**.624**	.130	-.328
V3TI-6	.276	**.534**	.075	.231

PCA Varimax with Kaiser Normalization.
*Rotation converged in 5 iterations.

Table 4.45 shows a clear factor-item distribution for Task S (Component 1) and Task I (Component 2), and very similar to Version 1, Task T is accounted for by two components. The first two items in Task T (V3TT-2 and V3TT-3) load on Component 4, the last two items (V3TT-5 and V3TT-6) load on Component 3, and the third item (V3TT-4) loads on both Component 3 and 4. When parallel analysis values were calculated, it was found that eigenvalues of only three factors exceeded the corresponding criterion values (see Table 4.46). Although the scree plot did not reveal a clear break after the third component (see Appendix 4.9), the PCA was repeated by constraining the number of factors to three. Rotated 3-component matrix is given in Table 4.47 in which loadings of the items on the components that are above .300 are marked in bold and the amount of variance accounted by each component is given in parentheses.

Table 4.46: Comparison of eigenvalues from PCA and criterion values from parallel analysis

Component Number	Actual Eigen Value from PCA	Criterion Value from Parallel Analysis	Decision
1	3.762	1.3881	accept
2	1.602	1.3058	accept
3	1.288	1.2397	accept
4	1.077	1.1903	reject

Table 4.47: Version 3 Rotated component matrix*
3 components retained

Item	**Component**		
	1 (23.51%)	**2 (10%)**	**3 (8.05%)**
V3TS-2	**.569**	.106	.165
V3TS-3	**.568**	.168	.018
V3TS-4	**.635**	-.005	.019
V3TS-5	**.561**	.086	.149
V3TS-6	**.533**	.000	-.061
V3TS-7	**.703**	.213	.112
V3TT-2	.075	.163	**.393**
V3TT-3	.280	.277	**.484**
V3TT-4	.111	.160	**.714**
V3TT-5	.044	-.081	**.710**
V3TT-6	-.057	.116	**.586**
V3TI-2	-.032	**.689**	.110
V3TI-3	.211	**.709**	.181
V3TI-4	.160	**.691**	.178
V3TI-5	.023	**.582**	-.001
V3TI-6	.286	**.543**	.187

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
*Rotation converged in 4 iterations.

PCA with three factors accounted for 41.57% of the variance in the data and yielded a clearer distribution of items to components; Component 1 accounted for Task S, Component 3 accounted for Task T and Component 2 accounted for Task I. To confirm the relation between the tasks and the components, a correlation analysis was run between the factor scores saved as variables and the total score of each task. Table 4.48 below gives the correlations between the factor scores and the total task scores. High correlation coefficients confirm the relation between the tasks and the components.

Table 4.48: Pearson task/component correlations on Version 3

Component	Task S	Task T	Task I
1	**.961***	.174*	.192*
2	.168*	.216*	**.948***
3	.127*	**.925***	.194*

*Correlation is significant at the .01 level (2-tailed)

In sum, PCA yielded a four-component structure for Version 1, a three-component structure for Version 2 and Version 3. In Version 1, Task T (text level) loaded on mainly two components, the first two items (V1TT-2 and V1TT-3) loading on a separate component on which no other items loaded. When these items were analysed, it was seen that the distractors were frequently chosen as the right answer instead of correct options in this task, a problem that was identified above. These items have less favourable item statistics (low CITC) and it is not surprising that they loaded on a separate component. In Version 2, Task T (text level) and Task I (intertextual level) loaded strongly on separate components; however Task S (sentence level) was problematic in that version and this was reflected in factor loadings of the task. Version 3 had a neat 3-component structure. It can be concluded that when the tasks did not include unwanted variance, Principal Component Analysis reflected the three-task design of the tests giving substantial support to the efficient operationalisation of careful reading at sentence, text and intertextual levels through repeated evidence across versions. It is also note-worthy that in parallel with item statistics, PCA confirmed Task I as a consistent task across versions.

4.2.2.5 Discussion on the Statistical Analyses of the Test Results

The analyses above showed that there were significant differences in the scores the test takers achieved in the tests designed to operationalise different types of careful reading. Therefore, it has been confirmed that tasks targeting different levels of careful reading can make differing demands on test takers. However, the expected difficulty continuum, the tasks at the sentence level (Task S) being the least difficult and the ones at intertextual level (Task I) being the most difficult was not confirmed in the data. Instead, the tasks at text level (Task T) proved to be most difficult and in one case Task S was more difficult than Task I. Before any conclusions can be made on this finding, two issues need probing; technical quality of the items and operationalisation of the reading processes in the tasks.

The first issue to be dealt with is whether the items worked well technically. A few problem instances in the tasks showed that the failure in attaining the expected results may be directly related to the technical aspects of the items which could have been improved had the test been piloted extensively before the administration, an issue which will be discussed as the major weakness of the study in the conclusion chapter. For example, the problems encountered in Version 2 Task S were easily amendable technical problems (e.g. word overlap between the first sentence of the paragraph and the incorrect heading) which nevertheless skipped the attention of the test developers and did not surface at the piloting stage. Although the defective items were excluded from the analysis, their impact on the whole task could not totally be removed as in a task where a test taker is supposed to choose correct headings among a list of alternatives, wrong choices at one point should affect the likelihood of correct responses in the rest of the task. However, despite the technical problems in sentence level tasks (Task S), taking Version 1 and Version 3, it can be claimed that sentence level tasks have been operationalised as intended, yielding the expected statistical results in terms of difficulty. Text level task (Task T) on the other hand, exerted a greater demand on the test takers, expectedly more than sentence level tasks (Task S), but unexpectedly more than intertextual level tasks (Task I). What surfaced in the technical analysis of Task T was that the distractors were too strong in some cases, and minor details in the distractors were not sufficiently paid attention to. Although several revisions had been made to the summary statements after piloting them (see Section 4.2.1), the complexity and implicitness of information in some summary statements might not have been resolved sufficiently, possibly making the task unnecessarily challenging. In the same manner, although several improvements had been made to make the instructions clearer, the researcher does not know whether the problem was successfully removed as she did not perform *a posteriori* verbal protocol analysis due to time limitations. The fact that some statements summarised one paragraph and others more than one paragraph did in fact prevent paragraph level processing and encouraged text level reading (see Section 4.2.1); however, if the instructions were still confusing despite the revisions, unwanted difficulty might have been added to the task.

Another issue that still needs addressing is whether it was feasible to mark the ordering of the summary statements as it was seen in several cases during marking that some test takers could choose the correct statements but could not put them in the right order (e.g. Version 2). As one test taker pointed out in the verbal protocol, the difficulty of the text level task (Task T) may well be due to the complexity in its design rather than the genuine difficulty of text level reading as opposed to intertextual reading. '*One sentence per paragraph would be easy. You have to read very carefully and you have to keep in mind the text. Remembering, going back to the text, checking this and checking that... You have to do so many things at the same time'*. Therefore, it is not possible to claim that there were no technical problems with Task T although statistical findings were within the acceptable ranges. It should also be noted that item independence was a factor also in this task; incorrect selection of one item in a list would possibly affect the performance of other items.

Following the points raised above, it is suggested that no firm conclusions based on text level reading as it is operationalised through Task T in this study should be made. The third task, Task I on the other hand, exhibited no technical problems and worked as a moderately difficult task, discriminating well between the test takers from all levels.

Secondly, it has to be questioned whether the operationalisation of the designated reading processes in the test tasks is sufficiently accurate. This is not a matter to be clarified by statistical analysis alone; however, with the help of verbal protocol data (see Section 4.2.1), it can be claimed that at sentence (Task S) and intertextual levels (Task I), the tests have been successful in operationalising the targeted reading processes in the relevant tasks. However, the same conclusion cannot be drawn for the text level task (Task T) due to the reasons mentioned above. This task did trigger text level careful reading encouraging the test takers to read and understand the text in its entirety and discouraged paragraph level matching; however as mentioned above, complexity of the summary statements, irregularity of summarisation (one-to-one and two-to-one) and possibly lingering difficulty in the instructions render it difficult to assign the difficulty of the task only to the processing level operationalised in it.

Besides, difficulty of a task after all does not reside only in the operationalisation of the skills in it but in a multitude of factors; text characteristics as well as wording of the items, reading processes it evokes as well as cognitive demand it puts on the test taker, and also to the degree of the interpretation it requires from the reader (implicitness). For example, consider the example summary statement given in Section 3.2.4.2: *'Attempting to manage human evolution may be seen too ambitious for the reason that we cannot fully foresee the outcomes yet.'*, and an easier version of it: *'Human germline engineering is risky.'* Obviously, these two statements would place differing cognitive demands on test takers. Some of these factors had been taken care of by extensive a priori investigations that fed into the test specifications developed for the tests, yet some other factors, for example, the explicitness of summary statements may not have been controlled to a desirable degree. On the other hand, the intertextual level task (Task I) did activate the reading operations designated in the test specifications. Verbal protocols (see Section 4.2.1) supported the view that the participants needed to read the main text first and then they identified semantically related mini-texts. They also commented on the type of semantic link between the relevant part of the main text and the mini-texts. Therefore, Task I required the readers to form a situation model based on the first text and a situation model based on the mini-text and form an intertextual proposition - an *intertext predicate* - (building a rhetorical relation, i.e., support vs. oppose, agree vs. disagree, etc.) between the two texts. The building of this relation is not similar to building rhetorical relations within a single text as it is not facilitated by any textual coherence, thereby assumably being more demanding. However, noting that all mini-texts contained *confirmatory* information in the form of exemplification and elaboration, building meaning relations between the texts and mental integration of information may have been less demanding as it would be in the case of integrating *contradictory* information. It should also be noted that the intertextual reading task in the tests did not require the test takers to form a full documents model as *source evaluation* was not operationalised as a part of the task. Further investigation of the cognitive processes activated by the task and a more comprehensive operationalisation of intertextual reading processes are needed. Despite these, it can be said that Task I successfully operationalised, as determined in the test specifications, the process of *building up semantic relations between the texts,* and this is the first but nevertheless an important step of the process of integrating information from multiple texts.

Otherwise, despite all the problems stated above, it should be underlined that the test study presented in this section has revealed quite satisfactory results as it is admittedly a difficult task to implant the same level of difficulty across three versions of a test at the test development stage without any help of item statistics. Item difficulty is usually controlled through statistical analysis and versions are equated through statistical manipulation at the pilot stage. This study lacked that opportunity; however, three fairly equable versions were produced and this was confirmed by ANOVA, which showed that the tasks are significantly different but the versions are not. Had Version 2 Task S been better designed, there would presumably be much less task-version interaction. Besides, Principal Component Analysis confirmed the congruence between the dimensions of the reading construct as reflected in the test specifications and the test itself. ANOVA and PCA results above strongly suggest that the tasks elicited different levels of performance and different patterns of response.

4.3 Research Question 3) Is cognitive processing different in the various types of careful reading?

The investigation of the third research question was done through the statistical analysis of the cognitive proforma data. Cognitive proforma was designed to elicit *reading operations* that a test taker might use for each task and *reading processing level*, i.e. how much of the text they needed to process to answer the question (see Table 4.49)

Table 4.49: Cognitive validity proforma

To find the answer to a question, I tried to ...		TASK S	TASK T	TASK I
Q1	match words that appeared in the question with exactly the same words in the text	▢	▢	▢
Q2	quickly match words that appeared in the question with similar or related words in the text	▢	▢	▢
Q3	read only certain sentences of the text slowly and carefully	▢	▢	▢
Q4	read the whole text slowly and carefully	▢	▢	▢
Q5	connect information from one text and compare with information in other texts	▢	▢	▢
Q6	read relevant parts of the text again	▢	▢	▢
Q7	use my knowledge of how texts like this are organised	▢	▢	▢
Q8	connect information from the text with knowledge I already have	▢	▢	▢
I found the answer...		**Task S**	**Task T**	**Task I**
Q9	within a single sentence	▢	▢	▢
Q10	by putting information together across sentences	▢	▢	▢
Q11	by understanding how information in the whole text fits together	▢	▢	▢
Q12	by understanding how information in two different texts fits together	▢	▢	▢
Q13	I knew the answer without reading the text (s)	▢	▢	▢
Q14	I could not answer the question	▢	▢	▢

4.3.1 Means and Mean Comparisons

Table 4.50 presents the results from reading operations and Table 4.51 gives the results of reading processing levels. The data analysis was done using non-parametric Friedman tests on each proforma question. As 14 comparisons were carried out on the same test takers, a Bonferroni correction was applied to give a *p* value of .0036 (.05/14). The results were significant on all proforma questions except for Q6.

Table 4.50: Friedman test on the Reading Operation items (N=931)

		Reading Operations							
		Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**
Mean	**Task S**	.36	.61	.37	.27	.25	.47	.23	.16
	Task T	.23	.44	.31	.48	.31	.53	.20	.13
	Task I	.22	.41	.27	.42	.50	.49	.17	.10
Friedman's Chi Square		73.70	112.92	26.19	107.58	158.18	8.75	15.87	16.48
p		**.000**	**.000**	**.000**	**.000**	**.000**	**.013**	**.000**	**.000**

Table 4.51: Friedman test on the Reading Processing Level items (N=931) (sig. at .01 level)

		Reading Processing Level					
		Q9	**Q10**	**Q11**	**Q12**	**Q13**	**Q14**
Mean	**Task S**	.33	.45	.44	.10	.10	.06
	Task T	.15	.46	.62	.13	.05	.14
	Task I	.11	.38	.41	.60	.02	.21
Friedman's Chi Square		172.87	18.25	109.08	669.22	58.86	110.68
p		**.000**	**.000**	**.000**	**.000**	**.000**	**.000**

Using the non-parametric Wilcoxon signed ranks test, planned comparisons were conducted between the three tasks on each proforma question except for Q6 (see Table 4.52). Since three tests were involved in each comparison, a Bonferroni correction was applied to the significance level to give an adjusted *p* value of .0167 (.05/3).

Table 4.52: Post Hoc, Wilcoxon signed ranks test N=931 [sig. at .0167 level]

	TASK S	**TASK S & T**		**TASK T**	**TASK T & I**		**TASK I**	**TASK S & I**		**TASK S**
	mean	**z**	***p***	**mean**	**z**	***p***	**mean**	**z**	***p***	**mean**
Q1	.358	-7.00	.000	.226	-.53	.598	.217	-7.12	.000	.358
Q2	.612	-7.96	.000	.442	-1.84	.066	.407	-9.42	.000	.612
Q3	.372	-3.00	.003	.309	-2.11	.035	.269	-4.82	.000	.372
Q4	.271	-10.22	.000	.478	-2.87	.004	.419	-7.02	.000	.271
Q5	.255	-3.24	.001	.312	-8.76	.000	.501	-11.21	.000	.255
Q7	.233	-1.93	.054	.202	-2.17	.030	.171	-3.82	.000	.233
Q8	.160	-2.17	.030	.128	-1.93	.054	.104	-3.94	.000	.160
Q9	.332	-9.22	.000	.152	-2.59	.010	.115	-11.51	.000	.332
Q10	.452	-.36	.720	.459	-4.01	.000	.380	-3.45	.000	.452
Q11	.439	-8.18	.000	.615	-9.54	.000	.406	-1.56	.119	.439
Q12	.102	-1.97	.049	.128	-18.99	.000	.604	-19.91	.000	.102
Q13	.101	-4.31	.000	.049	-3.33	.001	.024	-7.27	.000	.101
Q14	.063	-6.30	.000	.136	-4.86	.000	.208	-9.93	.000	.063

In comparing tasks in terms of reading operations and processing levels, each task will be compared to the other two tasks to reveal the reading operations and processing levels that are utilised significantly more than the other two tasks.

Task S versus Task T and Task I

There were three reading operations and one processing level that the test takers utilised significantly more in Task S than in Tasks T and Task I:

Q1: matching words that appeared in the question with exactly the same words in the text *(Task S: 35.8%, Task T: 22.6%, Task I: 21.7%)*

Q2: quickly matching words that appeared in the question with similar or related words in the text *(Task1: 61.2%, Task T: 44.2%, Task I: 40.7%)*

Q3: read only certain sentences of the text slowly and carefully *(Task S: 37.2%, Task T: 30.9%, Task I: 26.9%).*

Q9: finding the answer within a single sentence *(Task S: 33.2%, Task T: 15.2%, Task I: 11.5%).*

Q13: finding the answer I knew the answer without reading the text (s) *(Task S: 10.1%, Task T: 4.9%, Task I: 2.4%).*

Thus, Task S, which was designed to test careful reading at the sentence level, seems to require test takers to match words that appeared in the question with exactly the same or similar words in the text or to read only certain sentences of the text slowly and carefully, while perhaps processing only a single sentence, or perhaps without reading the text(s)t–

a result which may imply that the test takers did not read the text in its entirety, rather than that they ignored the text altogether.

Task T versus Task S and Task I

There was one reading operation and one processing level that the test takers utilised significantly more in Task T than in Task S and Task I:
Q4: reading the whole text slowly and carefully *(Task S:27.6 %, Task T:47.8 %, Task I:41.9 %).*
Q11: finding the answer by understanding how information in the whole text fits together *(Task S:43.9 %, Task T:61.5 %, Task I: 40.6%).*

This is in accordance with the test-designers' intention, as Task T was designed for creating a macrostructure for the given single text.

Task I versus Task S and Task T

When performing Task I, a significantly larger proportion of test takers, Q5: tried to connect information from one text and compare with information in other texts *(Task S: 25.5 %, Task T:31.2 %, Task I:50.1 %).* Furthermore, in Task I, significantly more test takers (60.4%) reported *Q12*: finding answers by understanding how information in two different text(s) fits together *(Task S:10.2 %, Task T: 12.8%, Task I:60.4 %).* These findings confirm that, as intended, test takers performed Task I by combining information across texts.

4.3.2 Discussion on the Cognitive Proforma Results

Although the mean differences were not sharply different in all cases, they were all significant and suggested that, as intended, each of the three tasks activated different cognitive processing demands on the test takers. It is surprising however, to see such cases where the test takers, though few in number, reported that they found the answer *by understanding how information in two different texts fits together* when there was only a single text, or how the test takers spotted the answer within a single sentence in Task T and Task I. This may suggest that there might be some ambiguity in the cognitive proforma items. Although it is always desirable to familiarise the participants with such tools before the administration, this was not possible in this case where approximately 1000 participants sat for an exam. The contention is that if the test takers had been familiarised with the proforma beforehand, the results would differentiate between the cognitive processes employed in each task more clearly. Nevertheless, the cognitive proforma study supported a conclusion that cognitive processing is different in various types of careful reading and the processes are in line with the specifications of careful reading skills in the tests.

4.4 General Comments on the Results

This chapter has presented six studies that have built up a theoretical argument claiming that careful reading in academic environments has three identifiable layers which are shaped by different reading purposes and manifested through different cognitive processes, therefore these three careful reading types need to be assessed in the tests of Academic English. The emphasis in this research is on building theoretical premises for the assessment of careful reading at intertextual level, which is the least researched skill area, and which has not yet reached EAP assessment.

At the initial stage with the questionnaire study, the present research established that reading multiple texts to integrate information is perceived as an important academic reading skill among the first and second year university students. Through the reading diary study, a more detailed understanding of the reading behaviours of students reading for assignment purposes has been gained. The study showed that reading purposes and processes requiring building informational links across texts were predominant. The interview study confirmed that careful reading processes of the participants were strongly intertextual in which several texts were connected and the situation described in each text was integrated in a high order documents model. The study confirmed the intertextual reading processes identified in the literature (ex: Perfetti et al, 1995; Britt and Sommers, 2004; Rouet, 2006) and highlighted several differences and additional evaluative strategies in the reading processes of the participants. The suggested frame of 'reading for writing purposes' summarised the prominent reading processes involved in reading at intertextual level. These studies established the importance of intertextual reading in academic environments and provided insight into the different cognitive processes that delimit the types of careful reading to inform the design and use of the language tests developed to be used in the study. The language tests were used to investigate whether test takers would perform differently on the tests of different types of careful reading, more specifically whether there were differences both in product and process in carrying out these careful reading activities in the tests. The product analysis was carried out through statistical methods and the processes were investigated through verbal protocols and cognitive validity proforma.

The three research questions formulated at the onset of the study were:

Research Question 1) What are the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution?

The finding is that university students engage in all three levels of careful reading: They focus on sentences to form propositional meaning in their mind and read incrementally up to whole text level when they think the text is important and when they read for genuine academic tasks, they read to combine information across texts.

Research Question 2) Do test takers score differently on tasks operationalising three careful reading types?

This research question has been answered positively yet only tentatively as the expected continuum of difficulty has not been confirmed through the use of the test tools developed for the study.

Research Question 3) Is cognitive processing different in the various types of careful reading?

Differences in the cognitive validity proforma data between the cognitive processes used for each task were found. This was in line with the findings from the verbal protocols conducted at the *a priori* stage.

Detailed discussions on each of these investigations are provided on the basis of the six studies performed to answer the above research questions. The next chapter will present the summary and implications of the findings, limitations of the study, and suggestions for future research.

CHAPTER 5

CONCLUSIONS

5.1 Introduction

The purpose behind all assessment procedures is to infer – based on the test scores – the extent to which a test taker would be able to perform the pertinent skill in real life situations. In designing EAP reading tests, we are interested in inferring the reading ability of a test taker in situations in which he or she would read academic texts and perform academic tasks. As stated previously, the first step in doing this is to base the test on a construct definition so that we can assess the extent to which the test results can generalise beyond the testing situation.

The motivation behind this study was to contribute to our understanding of the academic reading construct both in its cognitive and contextual aspects and investigate, where the purpose is to generalise test results to an academic environment, whether EAP reading tests need to assess careful reading not only at lower levels of processing (as is the case in many EAP reading tests) but also at higher levels including the intertextual level. The findings from the studies will be revisited below and implications of each study will be given in order in Sections 5.2, 5.3 and 5.4. Section 5.5 will present the implications of the study for EAP testing and teaching from a broader perspective. The following sections will present the limitations of the study and suggestions for further research.

5.2 Summary of the Results and Implications on Research Question 1: What are the different types of careful reading for academic purposes that undergraduates are faced with in a tertiary institution?

The study set out to establish the academic needs of students by referring to the students' own perceptions rather than the views of university faculty, and investigated in depth the related cognitive processes and contextual features of academic reading as the students were engaged in a genuine academic task.

The first research question was formulated and three studies – the questionnaire, the reading diary and the interviews – were designed to investigate the issue. The questionnaire study showed that students perceived integrating information across texts as an important skill and its importance increased from Year 1 to Year 2 students and from English as an additional language (EAL) group to native speakers. Establishing the importance of multiple texts reading from the perception of students was complementary to several studies that were based on faculty perceptions (e.g. Moore et al. 2010) and it showed that the importance of integrating information across texts from students' point of view may increase in relation to academic experience of the students (Carlson, 2001, Spivey and King, 1989) and in this case according to language background.

This may suggest that certain students at the beginning of their university study, especially the ones from non-native backgrounds may not have the understanding that meaning making in academic study, at least in the Western world, depends largely on combining information from multiple sources, in which taking a critical stance towards the sources is usually required (Goodman, 2004) and this lack of knowledge may disadvantage them with regards to their success in the first years of their study. A large number of EAL students are accepted at English medium universities partly on the basis of the scores they achieve in EAP reading tests assessing local reading comprehension which necessarily encourage local reading comprehension in the language learning and test preparation phase. Investing time and energy in studying for such tests, students may therefore be misguided in terms of the level of proficiency in reading necessary for university study.

The second study that investigated the first research question was the longitudinal diary study that provided data on the reading behaviours of a group of students reading for assignment purposes. The diary study helped identify the reading purposes and processes involved in the reading into writing process. Certain reading processes appeared to be activated by certain reading purposes; e.g. careful reading at different levels matched with sentence, text and intertextual reading purposes. The readers reported the use of text and intertextual level reading purposes predominantly and background knowledge surfaced as important in all kinds of reading. The findings confirmed the nature of academic reading as constructing meaning within and across texts (Spivey and King, 1994; Stromso and Braten, 2002). Frequently endorsed reading purposes such as reading texts carefully to remember major ideas, comparing and contrasting the information in the text, evaluating writers' ideas and comparing view points attested that academic reading involves reasoning with documents (Rouet et al., 1996) in a complex process of 'discourse synthesis' (Spivey and King, 1989). This process also involved searching for relevant information at the initial stages and note-taking, and as mentioned above required profound use of background knowledge of content and discourse organisation.

Goldman (2004) determines that successful learners use whatever content and discourse knowledge is available to them to manage the search for relevant information and the construction of relationships among ideas within a text and across texts. They not only form a mental model of the situation within a text but they are also able to organise representations across situations into a coherent whole. The responses in the diary study reflected similar attempts by the participants, providing us with further evidence on the essentiality of global careful reading skills in learning from texts. It is worth repeating that in academic settings learning is a complex process of integrating information across documents and integrating it with one's knowledge base through evaluative and elaborative reading (Britt and Sommer, 2004; Goldman and Rakestraw, 2000; Langer and Flihan, 2000) and the diary study attested to it.

The third step in the study on the nature of academic reading was a close-up investigation of the reading processes of a group of students preparing an assignment. The interview study manifested several important details on intertextual reading process. Searching for relevant information, text based reading, comparison of information across texts and evaluation of it with respect to what the participants had previously read, source evaluation and the evaluation of information in terms of its usability in the readers' evolving texts were predominant processes the participants used. The interview study supported several points put forward by research:

*Reading in academic settings is a purposeful, task-oriented, multi-faceted activity in which all bits of relevant information are meaningfully and systematically related to one another (Goldman, 2002).

*Topic knowledge, knowledge about texts and the relationship between texts and the knowledge of disciplinary conventions are important (Goldman and Bloome, 2004, Hartman, 1995; Perfetti et al., 1995; Spivey and King, 1989).

*Texts are connected through intertextual semantic links and situations described in texts are connected through in a higher order situations model (Rouet et al., 1996; Perfetti et al., 1996, Wineburg, 1991).

*Source characteristics are incorporated with the intertextual semantic links to give an intertext model (Perfetti et al., 1999; Rouet, 2006).

This study also revealed discipline-specific variations in the evaluation of source characteristics and the importance of usability judgements which have been discussed in greater detail in sections 4.1e3.6 and 4.1e4. When all the points discussed above were brought together, the reading operations, purposes, processes and outcomes were summarised as in Table 4.17 in Section 4.1.4., which gives a summary of reading for writing purposes.

The theoretical implication of the findings from the three studies summarised above is then that reading theories should account for intertextual reading. Following this, it can be suggested that the Khalifa and Weir (2009) reading framework used in this study should include a more elaborate component of a documents model that will reflect the important processes of intertextual reading. Therefore, it can be suggested here that an intertext model in which the processes of 'intertextual proposition formation' describing the rhetorical relations built between the texts and the evaluation of source characteristics should be included in the model. It should also include a 'situations model formation' process in which situations described in texts are integrated. These two processes should explain the construction of an organised representation of multiple texts (Perfetti et al., 1999). The knowledge base component in the framework should also include 'documents knowledge' as a part of background knowledge a reader should bring into reading at multiple texts level. Lower levels being the same, additions to the model can be as follows:

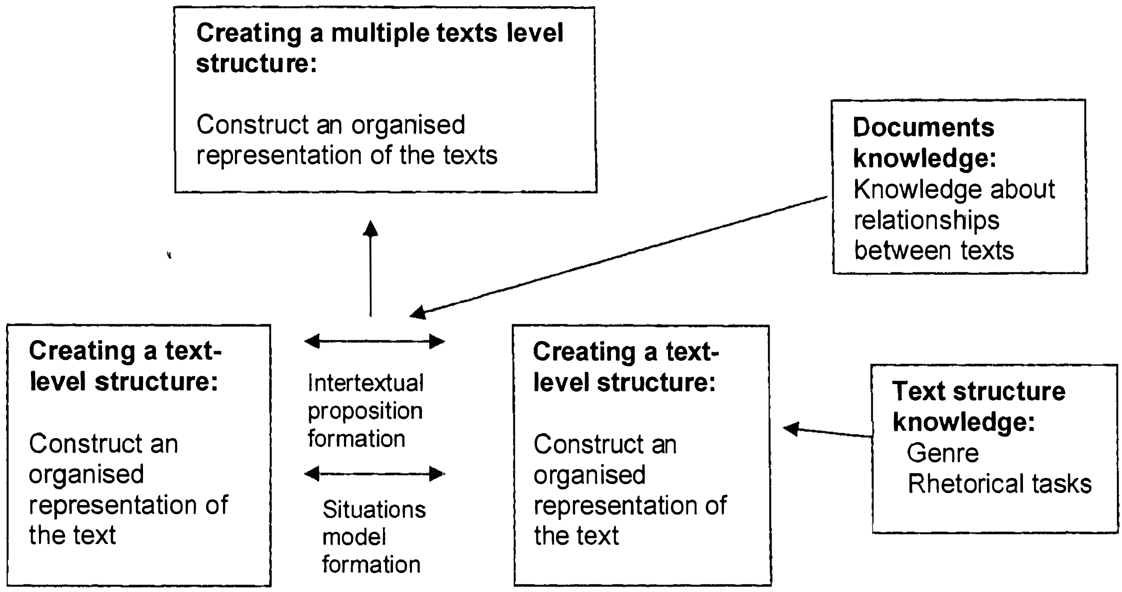


Figure 5.1: Suggested 'documents model' level for the Khalifa and Weir (2009) framework

Moreover, should the element of a 'writing task' as a governing factor in the reading process be included, then a higher level of 'creating an organised mental document/task' can be added. Similar multiple texts level reading processes can account for the integration of information into the evolving text (the task). However, documents model formation at text-to-task level should then be controlled by 'judgments of usability' in which the relevance of the information to the text being created is evaluated in the light of certain characteristics of information such as novelty and clarity. The judgements of usability are made taking into consideration the task purpose (i.e.: whether the information is relevant to the discussion the reader wants to develop in his or her own task, if so in what particular sense), the task organisation (i.e.: where in the reader's text the information can be used), the content of the reader's text so far (i.e.: whether the information adds something new). Judgments of usability can be taken as monitoring processes that guide the reading for writing process in terms of selecting and organising information mentally. Here, they are placed on the left-hand side of the framework, which accounts for reader's goal and monitoring processes. For more detailed discussion of the components placed in the model below, the reader is referred to Section 4.1.3. Figure 5.2 represents a possible model that focuses on only reading processes that may account for reading into writing skill.

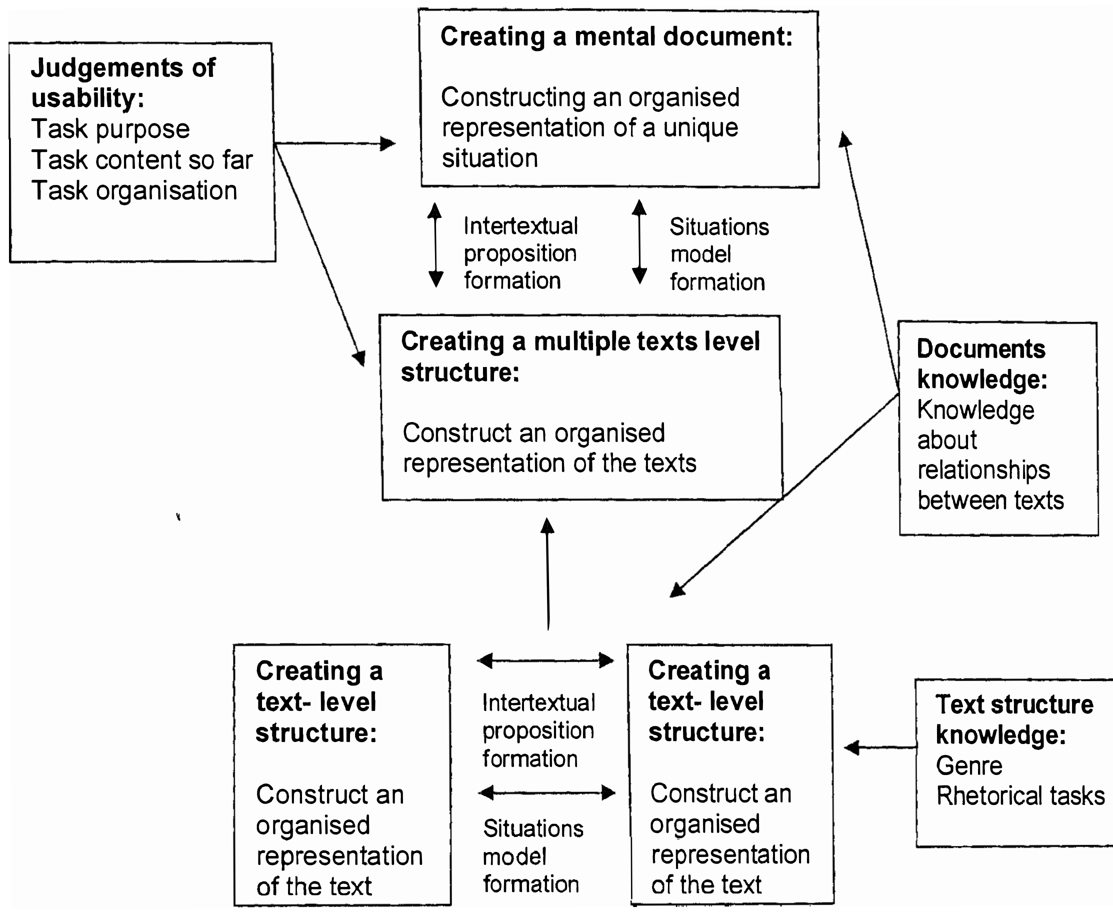


Figure 5.2: Suggested 'documents level reading into writing' model for the Khalifa and Weir (2009) framework

5.3 Summary of the Results and Implications on Research Question 2: Do test takers score differently on tasks operationalising three careful reading types at sentence, text and multiple text levels?

The investigation of the second research question exemplifies a theory-based test development. At the initial stage, where the bases for the assessment of careful reading skills were set, the need for a test that assesses careful reading skills at various levels was established. Analysis of the major English proficiency tests showed that reading across texts is either not assessed or assessed through written performance and usually scored as a part of writing sections of the tests. Secondly, the foundations for developing such a test were set through establishing contextual features and cognitive processes that would be operationalised in the test. Through a detailed review of the literature on textual features and investigation of web-based text analysis tools, a text analysis scheme, referred to as context validity proforma in the study, was developed. Cognitive processes that would be operationalised in the test were synthesised through the reading literature and the previous part of the study. These studies enabled theory-driven and research-based development of test specifications. Among the three tasks developed for the study, two (Task T and Task I) were innovative tasks, the former aiming to compensate the shortcomings of similar text level tasks, the latter being an unprecedented tool for assessing intertextual reading comprehension. The tests were developed in a focus group with iterative revision and verbal protocol study helped refine the tests and provided initial evidence for the congruence between the test specifications and the test tasks.

Statistical analyses of the test taker responses showed that there are differences between performances on the tests designed to assess three levels of careful reading. As suggested in the literature, local level reading proved to be the easiest task; however the expected difficulty continuum could not be confirmed on text and intertextual level tasks. Text level tasks seemed to activate recursive text level reading but it was also the most challenging task. The reason for this remained ambiguous; however clearer instructions, simpler summary statements and alternative scoring methods were suggested as improvements on the technical quality of the task, which otherwise is a promising alternative to compensate for the weaknesses of existing summary tasks (i.e. Cohen and Upton, 2006). Although there were no observed technical problems with the intertextual level task and the verbal protocols attested its intertextuality, the need to further investigate the processes it activates has been underlined. The need to operationalise in the task the intertextual processes beyond matching confirmatory information across texts was also emphasised. All in all, the second research question has been confirmed in the positive and this has certain implications for EAP testing.

Initially, it is suggested here that test tasks that putatively assess text level reading should be designed with care since tasks that require test takers to focus on smaller units of texts (i.e. paragraphs) first and then juxtapose information from such small units to form a summary frame do not necessarily operationalise the essential process of 'macrostructure' formation. As Cohen and Upton (2006) put it and the analysis in Section 3.2.1 showed, test takers may arrive at correct answers without having to process the texts in its entirety. Text level tasks should require test takers to delete redundant and unimportant information and to reduce the whole text level information to its gist by observing the order of the information and the relations between the parts of the text (Kintsch and van Dijk, 1978).

This study has exemplified a text level reading task that required recursive 'whole passage processing' where test takers had to put summary statements in the order of the passage choosing among statements that do not match the paragraphs on one-to-one basis, thus compensating for the deficiencies observed in similar tasks. Despite the problems with the task discussed earlier in the study, it showed that text level reading processes such as macrostructure and situation model formation may be better operationalised through the suggested format.

Secondly, the study also introduced a 'multiple texts' task, operationalising the initial step of documents level reading, yet failing to support the expectation that it should be more difficult to process multiple as against single texts. Although it was found that the multiple text tasks were in most cases more difficult than the test tasks targeting the sentence level, they were invariably easier than tasks targeting the whole single text level. The problem most probably lies in the format adopted in the text level task that required simultaneous processing of text information, organisation of information in the text and information in the summary sentences.

On the other hand, the expectation in the design of the 'multiple texts' task was that the test takers would have to process the main text in its entirety to be able to see the relations between it and the accompanying mini-texts and as a result the task would initially involve macrostructure formation and then additional processes of integrating information. The test takers attested in the verbal protocols that they had read the whole text and related the mini-texts to the main text as designated in the task specifications. However, the task as it was designed operationalised only certain processes in the documents model reading. Remembering that intertextual reading is a complex process involving both reading and writing related processes, we can say that the processes pertaining to reading only start off with the establishment of the 'relatedness' of pieces of information as it was operationalised in the test tasks of this study; however intertextual reading processes stretch well beyond that. Especially when the incoming information is confirmatory with the same set of relationships built between texts, the cognitive demand on the reader might be much lower than it is the case with conflicting propositions. Since conflicting propositions can hardly form a single coherent representation of a situation (Perfetti et al., 1999), processing confirmatory and conflicting relationships might exert differing demands on the reader.

On the other hand, it does not seem justifiable to claim that intertextual reading tests should attempt to operationalise such processes as source characteristics evaluation, yet tasks that assess the recognition and evaluation of several types of rhetorical relations between texts can be designed. Such tasks can involve 'a writing task and purpose' at mental level and operationalise relatively more authentic intertextual reading processes. A more difficult 'multiple texts' task might involve processing of both confirmatory and contradictory information and organisation of text parts in an essay outline. Comprehension questions requiring comparison of information across texts can also be designed.

5.4 Summary of the Results and Implications on Research Question 3: Is cognitive processing different in the various types of careful reading?

This relatively brief section of the study comprised the analysis of cognitive validity proforma data to confirm the cognitive processes used in responding the test tasks. Although the mean differences were not acutely different in all cases, significant differences were found to suggest that cognitive processing pertaining to each task generally involves different processes. The findings were in parallel to the findings of the verbal protocol study and the statistical findings from the test data, confirming that careful reading at sentence, text and multiple-texts levels require different processes and therefore are distinctly identifiable skills.

In sum, all the findings from the studies put together suggested that reading at lower levels is a necessary, yet insufficient skill and that reading across documents is a norm rather than an exception in academic situations. Different levels of reading are identifiable and each requires identifiably different cognitive processes which readers need to perform effectively for successful learning in academic settings.

5.5 Broader Implications for Testing and Teaching

If understanding and learning in academic environments depend on constructing meaning across multiple information sources, and learning from multiple texts lead to better reasoning and retention (Goldman, 2004; Goldman and Bloome, 2004) then the presence or lack of this skill is of crucial importance in judging the readiness of students for university study. Success in lower level comprehension does not necessarily entail attainment in higher level reading skills. We cannot infer from success at sentence level operations that a reader would be able to form a clear coherent understanding of all the major ideas in a text or that he or she would identify the necessary links between texts and pick out the relevant information for integration as required in successful university study. Therefore, information from tests which do not successfully operationalise and assess higher level reading skills is insufficient and potentially misleading for both the test takers and receiving institutions concerning an applicant's proficiency and readiness for study. To maximise the validity of the inferences that would be made based on test results, academic reading tests should assess the processes of both text level reading and intertextual level reading.

In certain studies, it is claimed that tests are limited in the extent to which they can simulate language use in target situation and they are designed to test 'readiness to enter' academic world rather than the mastery of the skills students are likely to need there (Taylor, 2007). Although it may not be realistic to expect a reading test to operationalise all domain specific skills, and for intending students to be endowed with all university level skills, reading at the text and intertextual levels are not peculiar to higher education. On the contrary, they are at the core of all meaningful educational processes and literate life. Without them many individuals in or outside academic environments would be unable to deal with the varied information sources and problem solving activities required to function in a knowledge society (Goldman, 2004). Therefore, 'readiness' in whatever sense of the word, cannot be evaluated through tests of limited, local comprehension when learners are expected to perform much more sophisticated tasks in the receiving institutions.

Moreover, as mentioned above, an important facet of a test's validity is the impact it creates: its consequential validity. Large scale, gate keeping tests shape learning processes, especially in the case of language learning. Therefore, it is test developers' duty to design tests that operationalise desirable skills in their tests if these are to be accepted as valid.

Language tests that are less than comprehensive will not create positive washback for learners on how to improve the various careful reading skills in the course of preparation for academic life either. The more fully a reading test samples the skills involved in what we might conceive as real academic reading, the more beneficial impact it will have on the learning and teaching process in general. If reading can be divided into subskills which the skilled reader is believed to use successfully and if these can be effectively operationalised in language tests, learners will be better prepared in the course of preparing for the test for the challenges and demands they will face in their academic life.

Understanding and identifying reading processes at the intertextual level is not an easy matter and as yet our understanding of them is incomplete. However, as the present study has shown, there are a number of processes encased in this complex literacy skill and it is extremely important to tease these out in order to inform focused practices that will assist learners in attaining documents level reading skills: skills which are otherwise not easily developed by learners on their own. Guiding students in understanding how texts may relate to each other and analysing the meaning relations across them, providing exercises on comparing and contrasting information from different sources and on juxtaposing confirmatory and confirmatory information taking into source characteristics as well might be benefical steps in teaching 'multiple-texts' reading.

Accordingly, the study has drawn attention to the necessity of better understanding the unique processes that govern the reading process of discourse synthesis (reading into writing). As mentioned above, there is some emphasis both in language teaching and testing on the production side of discourse synthesis (Grabe, 2001) – integrated reading-writing skills being introduced in large scale tests – but no emphasis is placed on either teaching or assessment of intertextual reading processes per se.

Acknowledging the value of written summaries or essays that require the integration of information across texts, it is suggested that designing efficient reading tasks at text and intertextual levels without recourse to writing is also important both for teaching and testing purposes as these are complex reading skills that may not develop readily on their own.

5.6 Limitations of the Study

As with many studies in the social sciences, this study is also limited by being carried out in a particular setting, with a certain group of participants and by the use of specific data collection tools. The qualitative parts of the study are in depth but involve a limited number of participants, the quantitative analyses may reveal only relatively superficial features. The use of triangulation of multiple methods in this study is an attempt to compensate for the weaknesses of each approach by combining them. However, a major shortcoming in this study has been the lack of a substantial pilot study of the reading tests used. Unfortunately time limitations did not permit an implementation of the full pilot administration of the tests, which had been planned.

Had it been possible to trial the tests on a sufficiently large group of test takers to allow for statistical analysis, and had there been enough time to redesign the tasks where necessary, the results of the test data might have been be much clearer.

Following this, another problem that needed to be addressed was the scoring procedures used for the tasks. At the stage of test data analysis, the items found to be defective were removed from the data analysis. However, although the defective items were excluded from the analysis, their impact on the whole task could not totally be removed. In tasks where test takers were supposed to choose the correct headings from a list of alternatives, wrong choices at one point must have affected the likelihood of correct responses in the rest of the task.

This in fact opens up another important discussion on whether assuming total independence between the items in such tasks is a valid assumption or not. Wainer's (2007) newly introduced *Testlet Response Theory* offers a promising approach to analysis; however, in its present form it is beyond the scope of a project of this kind. Moreover, although item interdependence is an important issue that has to be taken account of, it should be pointed out that the present study used statistical techniques that are widely used for tests that employ similar matching tasks and therefore, the results should be commensurate with current practice in the field.

Finally, in retrospect, during the test data collection phase of the study, the test takers should have been familiarised with the cognitive validity proforma before they were asked to fill it in as the results suggested that there might have been some ambiguity in the proforma items. Had this been recognised earlier on the basis of a large-scale pilot, the results from the cognitive proforma study could have been much clearer. However, as mentioned above in relation to the test itself, this was not possible within the parameters of the project.

5.7 Suggestions for Future Research

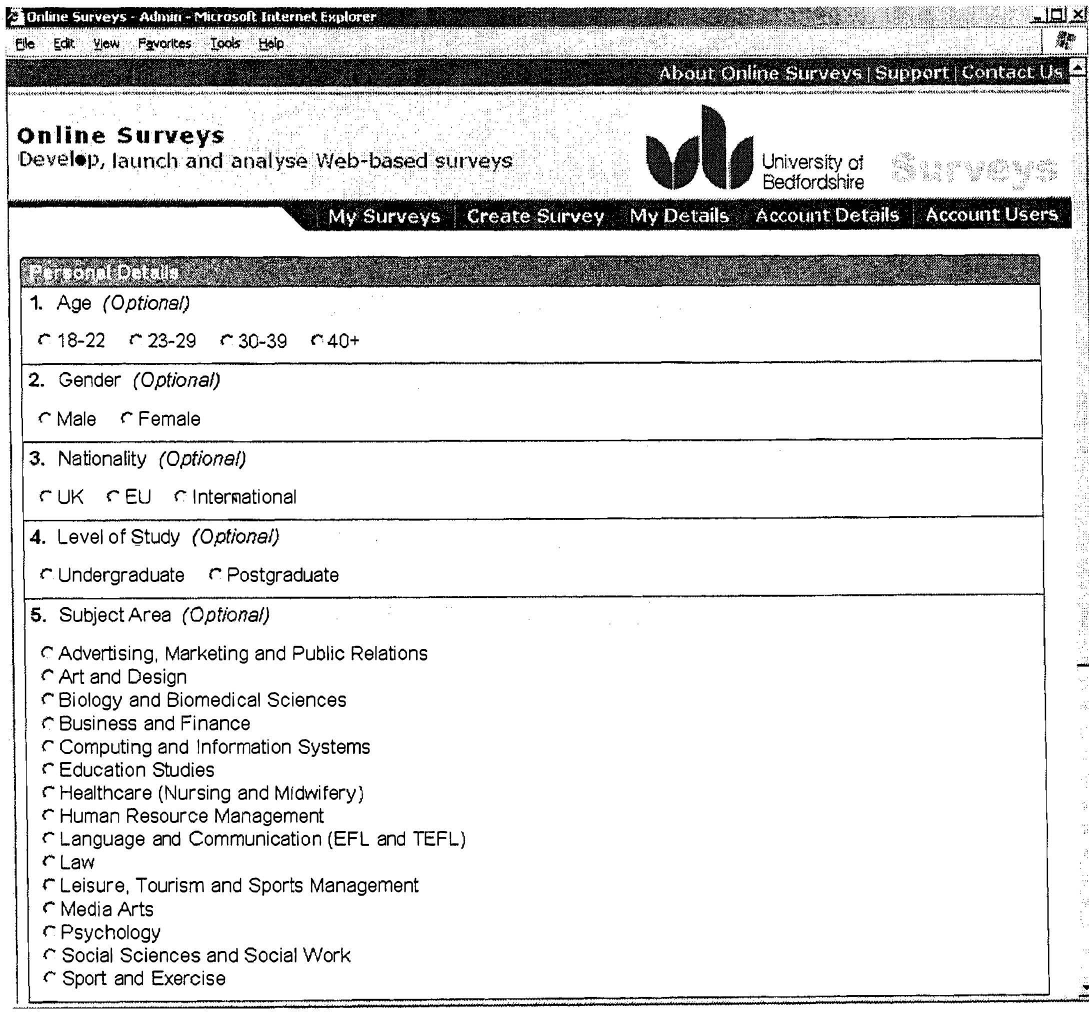
As an emergent area of interest, although there are studies that look into aspects of the documents model, there is a clear need for more detailed and extensive investigation in this area. Reading across texts should be investigated with respect to different academic disciplines, different academic tasks, different types of documents, different levels and types of readers. Strategies of multiple text reading should be compared between native and foreign languages. For a clearer understanding of the mechanisms of the process, longitudinal observations should be carried out with readers reading for genuine academic purposes.

Documents model reading should be deconstructed in order to identify the unique cognitive processes that shape each stage and differences should be identified between forming an understanding of a single versus multiple texts in more detail.

Experimental test tasks that attempt more exhaustively to operationalise the cognitive processes involved in reading across texts should be designed and verified through studies on test taker performance and faculty perceptions. Obviously, more knowledge needs to accumulate in the field and this needs coordinated and combined research efforts rather than individual and unconnected studies. As the knowledge on reading across texts accumulates, we will be better able to conceptualise and integrate it into the practices of teaching and assessment for more valid practices.

APPENDIX 3.1

Web-Based Questionnaire



Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

About Online Surveys | Support | Contact Us

Online Surveys
Develop, launch and analyse Web-based surveys

University of Bedfordshire

Surveys

My Surveys | Create Survey | My Details | Account Details | Account Users

Personal Details

1. Age *(Optional)*

- 18-22
- 23-29
- 30-39
- 40+

2. Gender *(Optional)*

- Male
- Female

3. Nationality *(Optional)*

- UK
- EU
- International

4. Level of Study *(Optional)*

- Undergraduate
- Postgraduate

5. Subject Area *(Optional)*

- Advertising, Marketing and Public Relations
- Art and Design
- Biology and Biomedical Sciences
- Business and Finance
- Computing and Information Systems
- Education Studies
- Healthcare (Nursing and Midwifery)
- Human Resource Management
- Language and Communication (EFL and TEFL)
- Law
- Leisure, Tourism and Sports Management
- Media Arts
- Psychology
- Social Sciences and Social Work
- Sport and Exercise

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

About Online Surveys | Support | Contact Us

Online Surveys
Develop, launch and analyse Web-based surveys

University of Bedfordshire Surveys

My Surveys | Create Survey | My Details | Account Details | Account Users

Academic Reading Research Project

For each statement below, show the extent of your agreement or disagreement:

5 Definitely agree
4 Mostly agree
3 Neither agree nor disagree
2 Mostly disagree
1 Definitely disagree

The following sources of information are important on my course.

9. Books

Select an answer

10. Journal Articles

Select an answer

11. Reports

Select an answer

12. Internet sites

Select an answer

13. Newspapers

Select an answer

14. Magazines

Select an answer

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

The following purposes for reading are important on my course

15. Searching texts to find information I can use in assignments and/or examinations *(Optional)*

Select an answer

16. Basic comprehension of just the main idea(s) in a text *(Optional)*

Select an answer

17. Understanding the meaning of the text as a whole: working out how the main ideas and details in a text relate to each other and to the author's purpose *(Optional)*

Select an answer

18. Integrating information from different texts for use in assignments and /or examinations *(Optional)*

Select an answer

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

How I read for assignments

19. I think carefully to ensure that I know exactly what I will be looking for before I start reading *(Optional)*

Select an answer

20. I quickly look through the whole of a text for a general understanding before doing anything else *(Optional)*

Select an answer

21. I gradually understand what a text is about by reading the sentences slowly and carefully in the order they occur *(Optional)*

Select an answer

22. I remember where relevant information is or mark its location for later use in writing my assignment *(Optional)*

Select an answer

23. I think of key words and quickly look for them or words with similar meanings to check if text is worth reading more carefully *(Optional)*

Select an answer

24. I look at the titles or headings of a text before deciding to read it carefully *(Optional)*

Select an answer

25. I first get an overall meaning of the text for example by reading the first paragraph and the conclusion, and the first sentence of the other paragraphs *(Optional)*

Select an answer

26. If I do not know the meaning of a word in a text, I try to work out its meaning *(Optional)*

Select an answer

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

27. I read a text slowly all the way through even if some parts do not seem relevant to my assignment *(Optional)*

Select an answer

28. I read slowly only those sections of a text I have marked as relevant when going through it quickly before *(Optional)*

Select an answer

29. While reading I try to relate content to what I know already and judge its valuee*(Optional)*

Select an answer

30. I look back at previous parts of the text to check meaning *(Optional)*

Select an answer

31. I try to understand how the text is organized: how the ideas and details connect with each other *(Optional)*

Select an answer

32. I make notes on relevant points from the text as I go along *(Optional)*

Select an answer

33. I integrate information from the text I am reading with information from other texts I have already read *(Optional)*

Select an answer

34. I read critically to establish and evaluate the author's position on a particular topic *(Optional)*

Select an answer

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

When I read for assignments I have difficulty with:

35. the time available to do the necessary reading. *(Optional)*

Select an answer

36. reading texts where the subject matter is complicated. *(Optional)*

Select an answer

37. words I do not know. *(Optional)*

Select an answer

38. sentence structures *(Optional)*

Select an answer

39. finding relevant information quickly *(Optional)*

Select an answer

40. lengthy texts *(Optional)*

Select an answer

41. lack of background knowledge to understand the content *(Optional)*

Select an answer

Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

42. making notes on information I will need *(Optional)*

Select an answer

43. reading carefully to understand the main ideas *(Optional)*

Select an answer

44. summarizing ideas from a text in my own words *(Optional)*

Select an answer

45. understanding a detailed logical argument *(Optional)*

Select an answer

46. reading critically to establish and evaluate the author's position on a particular topic *(Optional)*

Select an answer

47. relating the content of a text to my existing knowledge *(Optional)*

Select an answer

48. deciding what is important for me and what is not *(Optional)*

Select an answer

49. reading a text quickly to decide whether I should study it carefully *(Optional)*

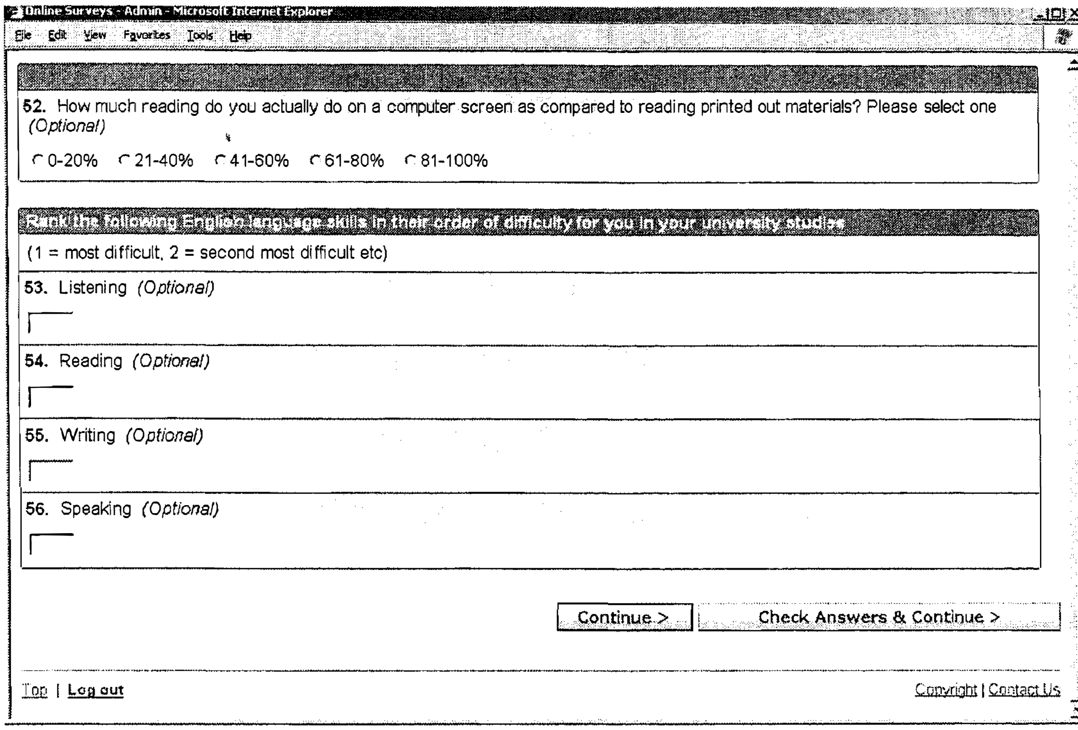
Select an answer

50. understanding the text as a whole; how main ideas and details are connected to each other *(Optional)*

Select an answer

51. integrating information from the text I am reading with information from other texts I have already read *(Optional)*

Select an answer


Online Surveys - Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

52. How much reading do you actually do on a computer screen as compared to reading printed out materials? Please select one *(Optional)*

0-20% 21-40% 41-60% 61-80% 81-100%

Rank the following English language skills in their order of difficulty for you in your university studies

(1 = most difficult, 2 = second most difficult etc)

53. Listening *(Optional)*

54. Reading *(Optional)*

55. Writing *(Optional)*

56. Speaking *(Optional)*

Continue > | Check Answers & Continue >

Top | Log out

Copyright | Contact Us

APPENDIX 3. 2

Students' Academic Reading Diary Form

Your name: ______________________ Email: ______________ Date: ______________

Section 1. Please answer about a text that you read for your assignment or dissertation

Title and main author: ____

This text was:

☐ A book ☐ An article in a journal ☐ A report

☐ An internet site ☐ An article in a newspaper or magazine ☐ Other ______________________

I read about ______ pages of the text, spending about ______ hours and ______ minutes.

On this occasion, I read the text or parts of the text...

☐ once ☐ twice ☐ three times ☐ four times ☐ more than four times

I read it

☐ on a screen ☐ on paper

I read it...

☐ at home ☐ in the library or resources room ☐ elsewhere ______________________

Section 2. About what you wanted from the text

When reading this text, I wanted to...

		Definitely agree	Mostly agree	Neither agree nor disagree	Mostly disagree	Definitely disagree
1	find specific words, phrases or figures	5	4	3	2	1
2	locate parts of the text that would be worth reading later	5	4	3	2	1
3	find support for or opposition to an idea I had in mind	5	4	3	2	1
4	find specific information to answer a question	5	4	3	2	1
5	get a basic understanding of just the main idea(s) in the text	5	4	3	2	1
6	decide whether the text would be worth reading	5	4	3	2	1
7	check my comprehension of a specific part of the text	5	4	3	2	1
8	understand and remember as much information as possible from the text as a whole	5	4	3	2	1
9	understand and remember as much information as possible, but only from relevant parts of the text	5	4	3	2	1
10	form an understanding of how the main ideas in the text relate to each other	5	4	3	2	1
11	evaluate the writer's ideas on the topic	5	4	3	2	1
12	compare the writer's point of view with that of other writers	5	4	3	2	1
13	judge how this text was related to other texts I had read in terms of supporting or developing ideas I had found elsewhere	5	4	3	2	1

Did you want to get any other kinds of information or ideas from the text?

Section 3. About how you read the text

		Definitely agree	Mostly agree	Neither agree nor disagree	Mostly disagree	Definitely disagree
1	I used the headings, titles, contents page or index to locate information quickly	5	4	3	2	1
2	I looked quickly for words relating to the topic of my assignment	5	4	3	2	1
3	I quickly matched words that appeared in the question with the same words in the text	5	4	3	2	1
4	I read key parts of the text such as the abstract, introduction and conclusion	5	4	3	2	1
5	I read slowly and carefully only certain sections of the text that I had decided were relevant to my needs	5	4	3	2	1
6	I looked for parts of the text the writer indicates to be important	5	4	3	2	1
7	As I read, I tried to connect information from different sentences in the text	5	4	3	2	1
8	I read the text slowly all the way through from beginning to end	5	4	3	2	1
9	As I read, I tried to form a summary of the ideas in my mind	5	4	3	2	1

		Definitely agree	Mostly agree	Neither agree nor disagree	Mostly disagree	Definitely disagree
10	I tried to combine information from this text with information from other texts	5	4	3	2	1
11	I looked at other texts while I was reading this one	5	4	3	2	1
12	I read the relevant parts of the text again	5	4	3	2	1
13	I used my knowledge of how texts like this are organised to find parts to focus on	5	4	3	2	1
14	I used my knowledge of the topic to help me to understand the text	5	4	3	2	1
15	I used my knowledge of the world to help me to understand the text	5	4	3	2	1

Did you read the text in any other way?

Section 4. Storing information from the text

		Definitely agree	Mostly agree	Neither agree nor disagree	Mostly disagree	Definitely disagree
1	I remembered where relevant information was in the text	5	4	3	2	1
2	I made notes on relevant points from the text as I went along	5	4	3	2	1
3	I highlighted parts of the text with a marker (or on screen highlighting tool) as I read	5	4	3	2	1
4	I copied and pasted (or copied down) relevant parts of the text as I read	5	4	3	2	1
5	I transformed ideas from the text into a draft of my assignment as I read	5	4	3	2	1

Did you store information from the text in any other way?

Section 5. Difficulties and usefulness

I found the text difficult to read in terms of...

		Definitely agree	Mostly agree	Neither agree nor disagree	Mostly disagree	Definitely disagree
1	The time available to do the reading	5	4	3	2	1
2	The subject matter	5	4	3	2	1
3	The vocabulary	5	4	3	2	1
4	The grammar	5	4	3	2	1
5	The organisation of the ideas	5	4	3	2	1
6	The length of the text	5	4	3	2	1

Please note anything else that caused you difficulties:

6	The text was useful for my needs	5	4	3	2	1

Please comment on reasons why the text was (not) useful.

APPENDIX 3.3

Interview Focus Questions

Procedure for think aloud protocol

The think aloud procedure is explained:
This exercise aims to identify the thought processes you use while you are reading a text of your choice for the purpose of preparing an assignment.
As you read, you are expected to verbalise what goes on in your mind, explaining how you make sense of what you are reading.
As you read, please try to answer the questions below in as much detail as possible.

Text selection:
How did you decide this article was worth reading?
What did you do just before you started reading?

While reading:
As you read along, what thoughts form in your mind?
What is the new information in the passage?
Have you learned anything new?
What is interesting in the article?

Linking with previous knowledge:
What do you remember from your previous readings about the issues discussed in the passage?
How has your accumulated knowledge helped you to understand this text?
How similar or different is the information in this text to what you have read previously?
How has this reading changed what you know?

Evaluation:
How do you evaluate the content of this text in terms of your writing goals? Is it useful?

APPENDIX 3.4

Context Validity Proforma

	quantitative parameters	qualitative parameters
Length	Number of words	
Vocabulary	Average character per word Standardised type-token ratio Lexical density 1000 word frequency 2000 word frequency 3000 word frequency Frequency<15K Academic Word List	
Grammar	Average words per sentence Average sentences per paragraph Flesch Kincaid reading ease Flesch Kincaid grade level Coh-Metrix readability	
Cohesion	Content word overlap (Coh-Metrix index 13) LSA Sentences all combinations mean (Coh-Metrix index 15)	
Rhetorical organisation		Does the text have an explicit organisational structure? (explicit) 1 2 3 4 5 (not explicit)
Genre		Identify the most appropriate category. 1. **text book** 2. magazine/newspaper article 3. research/academic journal article 4. report
Rhetorical task		Identify the most appropriate category. 1. exposition 2. argumentation/persuasion/evaluation 3. historical biographical/autobiographical/narrative
Subject specificity		Is the topic of the text of general interest or does it require subject specific knowledge on the part of the reader? (general) 1 2 3 4 5 (specific)
Cultural specificity		Is the topic of the text culture-neutral or is it loaded with specific cultural content? (culture neutral) 1 2 3 4 5 (culture specific)
Text abstractness	Concreteness of content words (Coh-Metrix index 44)	
Subject area		Mark as it applies. 1. Medicine & dentistry 2. Subjects allied to medicine 3. Biological sciences 4. Veterinary science 5. Agriculture & related subjects 6. Physical sciences 7. Mathematical sciences 8. Computer science 9. Engineering & technology 10. Architecture, building & planning 11. Social studies 12. Law 13. Business & administrative studies 14. Mass communications & documentation 15. Languages 16. Historical & philosophical studies 17. Creative arts & design 18. Education

APPENDIX 3.5

Test Specifications

CONTEXTUAL

Response Method: Multiple matching

Text length: 750-1000 words

Discourse mode

Genre: textbook, magazine/newspaper article, research/academic non-specialist journal article,
Rhetorical task: expository, argumentative
Pattern of exposition: may involve any
Explicitness of text structure: the reader should be able to cope with less explicitly organised texts
Content word overlap: 0.10
LSA, sentences all combinations mean: 0.26

Lexical and structural resources

Average character per wordt 5.14
Standardised type token ratio: 51-52
Lexical density: 0.56
1000 word frequency BNC %: 75-80%
2000 word frequency BNC %: 11-12%
3000 word frequency BNC %: 2-3%
Frequency< 15 K: 1%
AWL level: 10%
Average words per sentence: 21-22
Average sentences per paragraph: 3-4
Flesh Reading Ease score (0-100): 35-40
Flesh-Kincaid Grade level: 13-14
Coh-Metrix readability: 12-13

Nature of information: may involve abstract content
Concreteness, mean for content words: 355

Content knowledge: may require the understanding of specialised, academic content
Cultural knowledge: does not require any culture specific background

COGNITIVE

Careful reading at sentence level
Reading purpose: to comprehend meaning formed with in a sentence; understanding words, pronouns, and syntactic units within a sentence.

Cognitive processes: word recognition (accessing orthography, phonology and morphology of the words), lexical access (accessing the meaning of the words in a mental dictionary), syntactic parsing (analysing phrases and clauses at syntactic level), establishing meaning (forming a proposition).
Processing level: Local, sentence level

Careful reading at text level
Reading purpose: to comprehend overall information content of a text, to form a unified understanding (a mental summary) of a text
Cognitive processes: proposition formation, organisation of propositions, establishing meaning between propositions by integrating new information, building a mental model (macrostructure at discourse level), building a situational model that incorporates knowledge base.
Processing level: Global, text level

Careful reading at intertextual level
Reading purpose: integrating informational content of a text into the wider context of the topic
Cognitive processes: detecting and understanding the semantic links between the sources and identifying the nature of those links, integrating pieces of information into a new mental model
Processing level: Global, intertextual

APPENDIX 3.6

Cognitive Validity Proforma

Tick (✓) any sentences that describe what you did when you answered the questions in each part (**Task S, Task T, Task I**) of the test. You may tick more than one sentence for each part of the test.

To find the answer to a question, I tried to ...

		Task S	Task T	Task I
1	match words that appeared in the question with exactly the same words in the text	☐	☐	☐
2	quickly match words that appeared in the question with similar or related words in the text	☐	☐	☐
3	read only certain sentences of the text slowly and carefully	☐	☐	☐
4	read the whole text slowly and carefully	☐	☐	☐
5	connect information from one text and compare with information in other texts	☐	☐	☐
6	read relevant parts of the text again	☐	☐	☐
7	use my knowledge of how texts like this are organised	☐	☐	☐
8	connect information from the text with knowledge I already have	☐	☐	☐

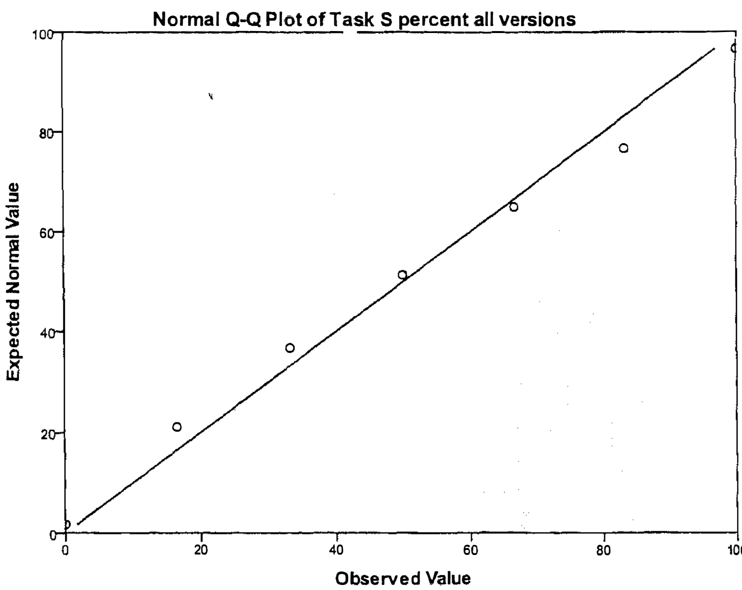
I found the answer...

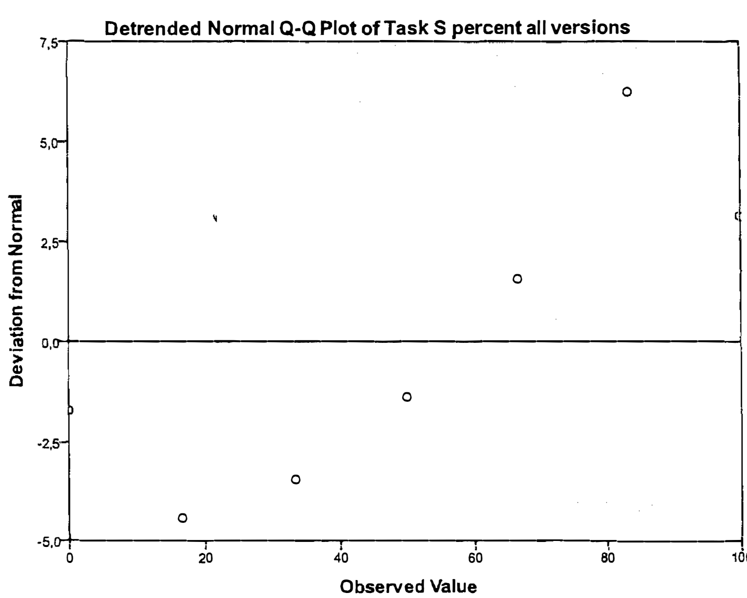
		Task S	Task T	Task I
9	within a single sentence	☐	☐	☐
10	by putting information together across sentences	☐	☐	☐
11	by understanding how information in the whole text fits together	☐	☐	☐
12	by understanding how information in two different texts fits together	☐	☐	☐
13	I knew the answer without reading the text (s)	☐	☐	☐
14	I could not answer the question	☐	☐	☐

Appendix 4.1

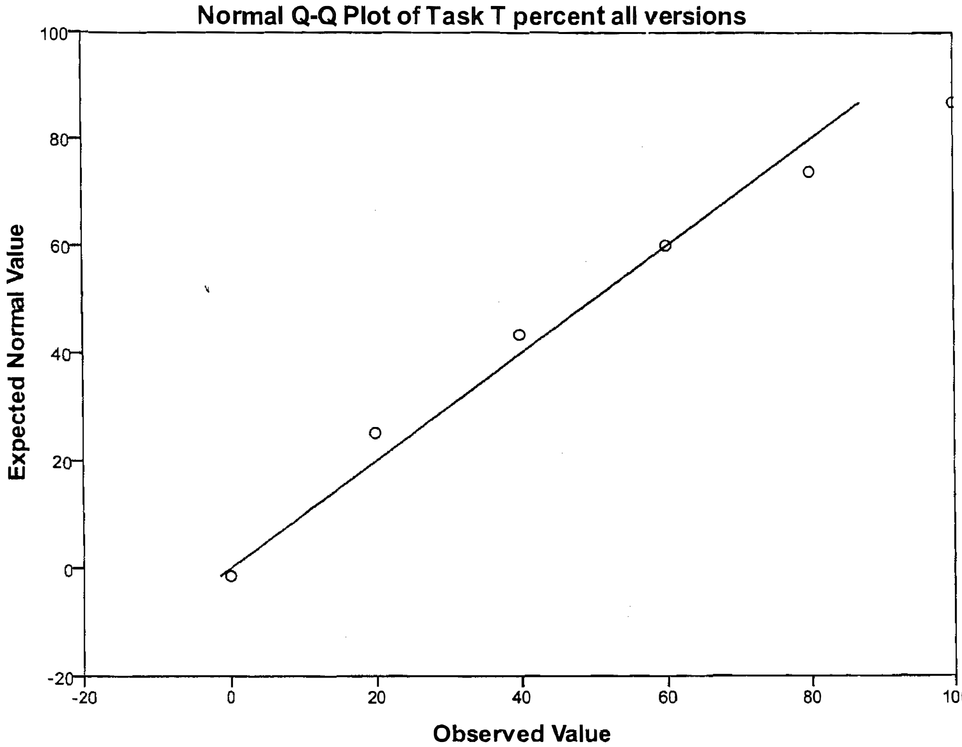
Normal Q-Q plots
All versions

TaskaS

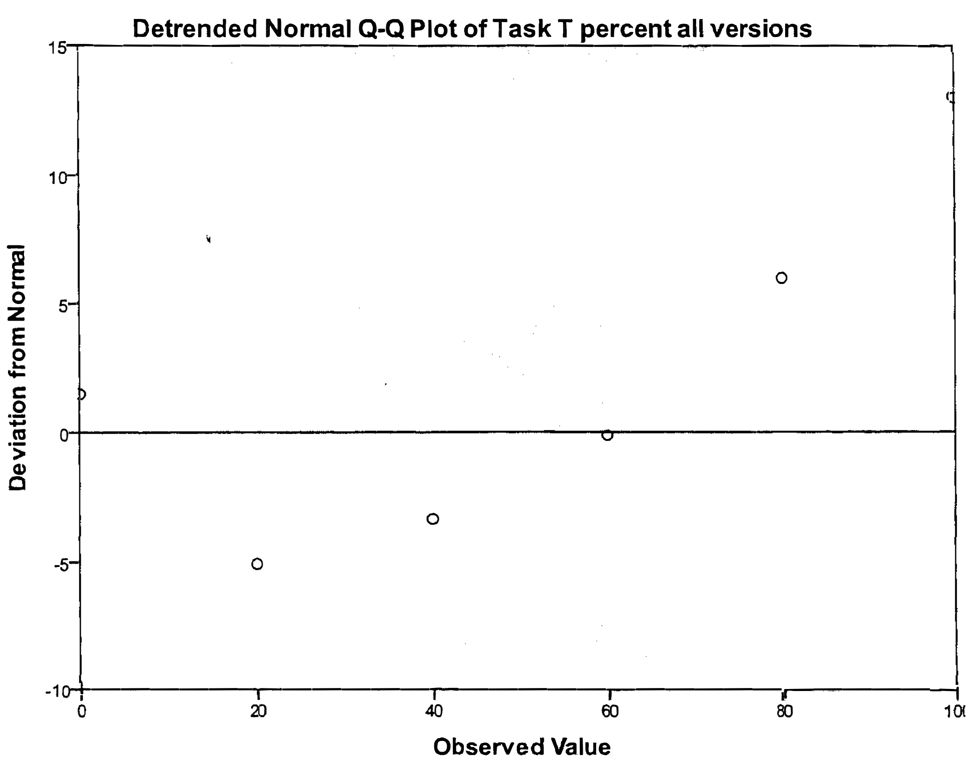




TasksT

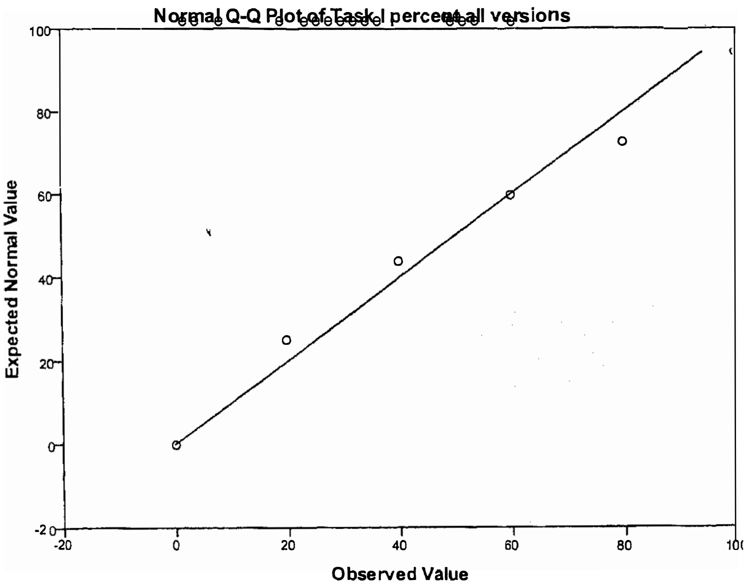


Normal Q-Q Plot of Task T percent all versions
Expected Normal Value
Observed Value

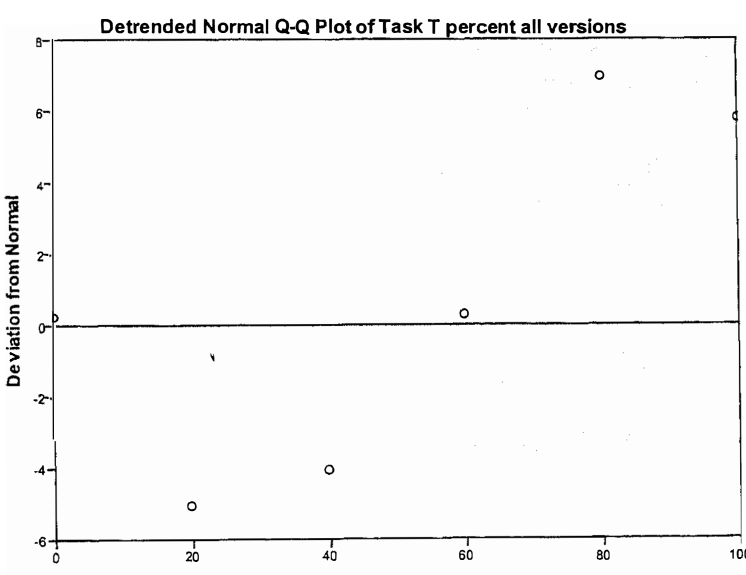


Detrended Normal Q-Q Plot of Task T percent all versions
Deviation from Normal
Observed Value

Task I



Normal Q-Q Plot of Task I percent all versions
Expected Normal Value
Observed Value
100
80
60
40
20
0
-20
-20
0
20
40
60
80

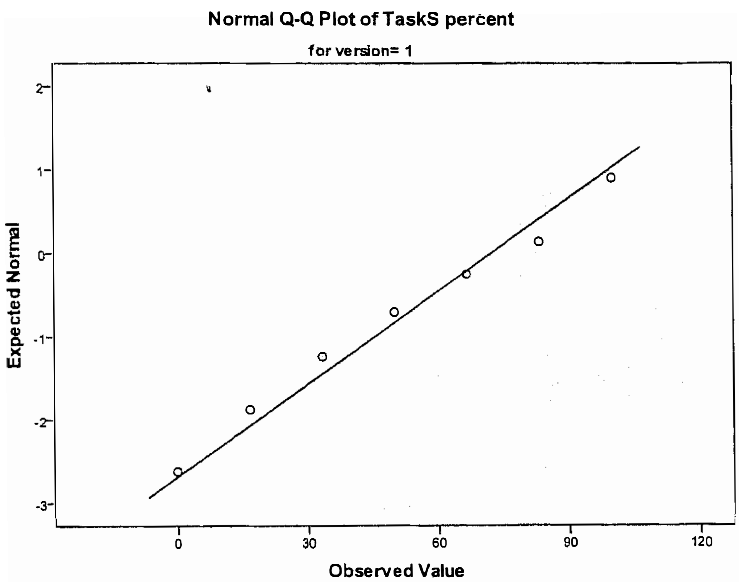


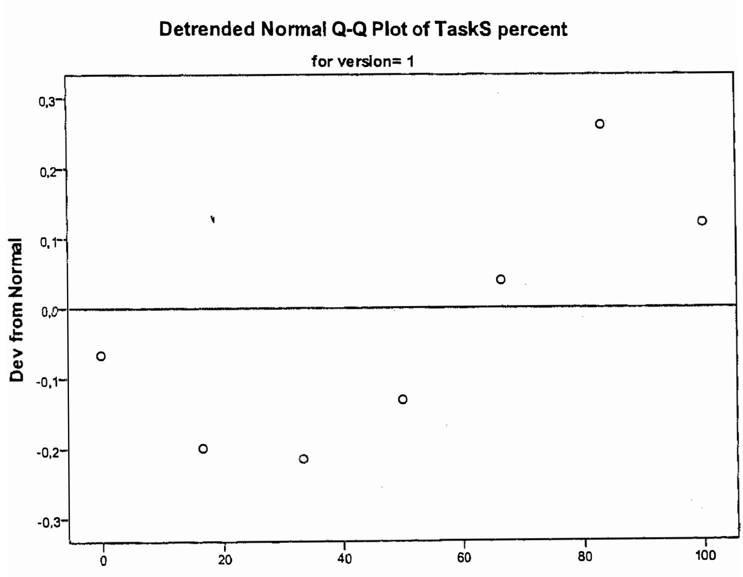
Detrended Normal Q-Q Plot of Task T percent all versions
Deviation from Normal
Observed Value
8
6
4
2
0
-2
-4
-6
0
20
40
60
80

Appendix 4.2

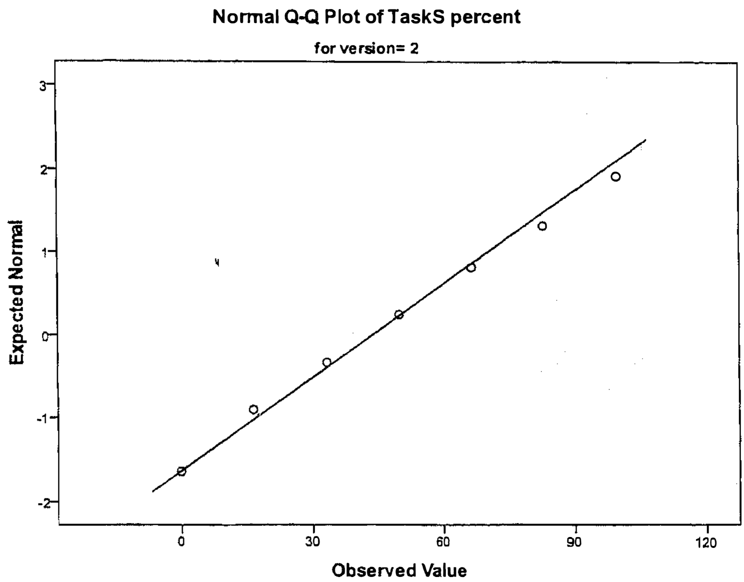
Normal Q-Q plots: By version

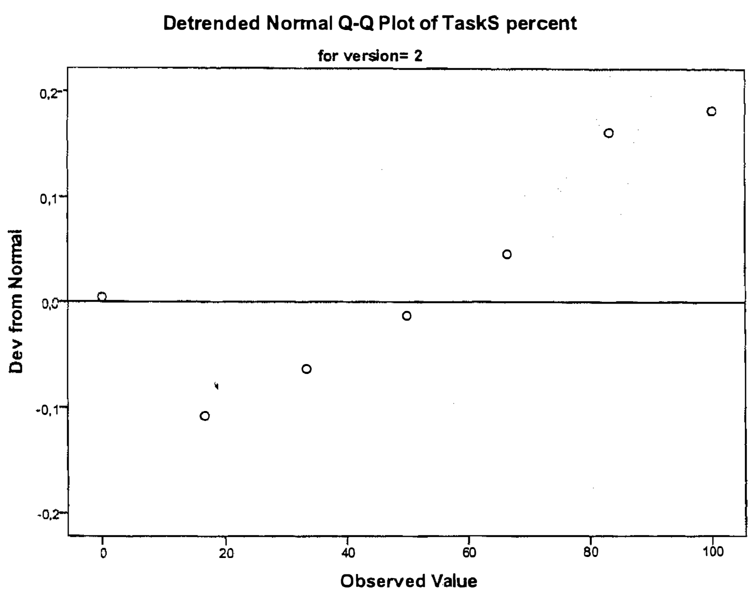
Version 1 Task S



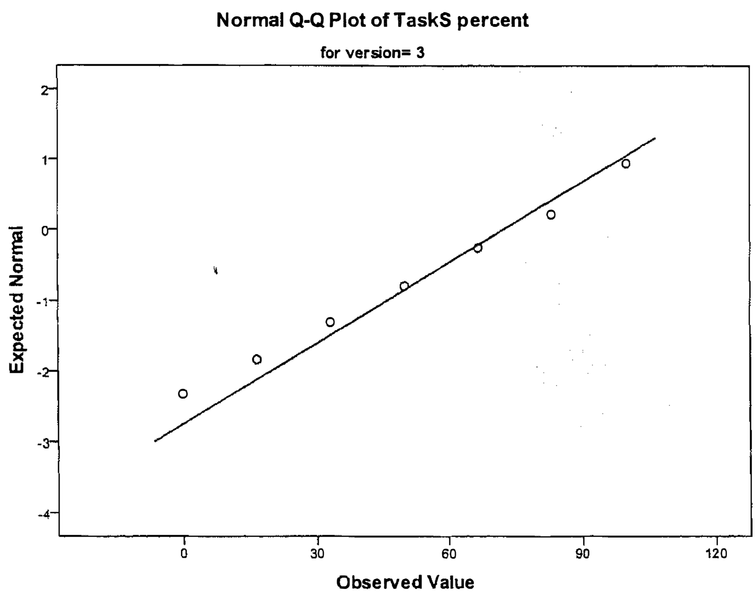


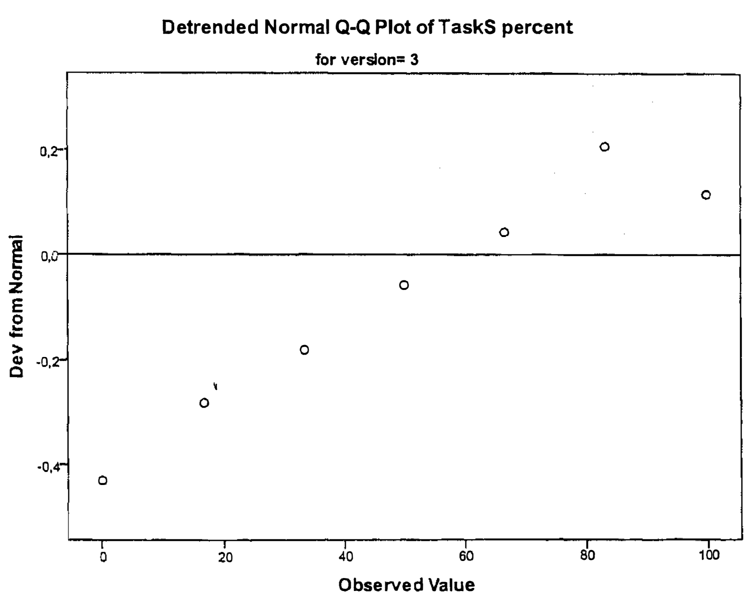
Version 2 Task S

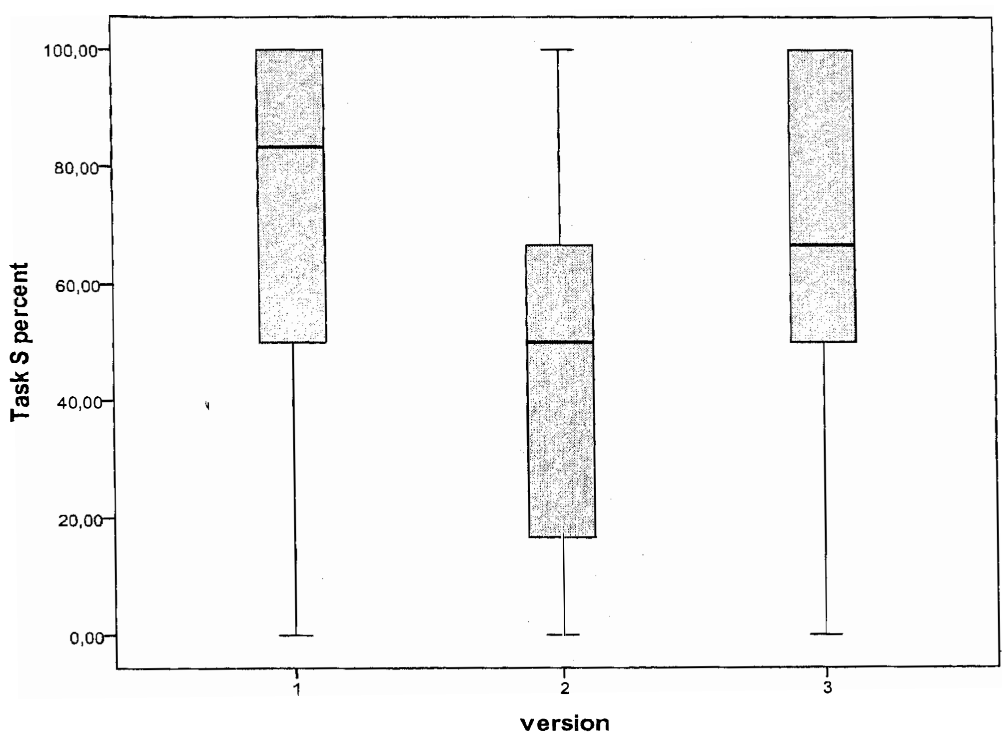




Version 3 Task S

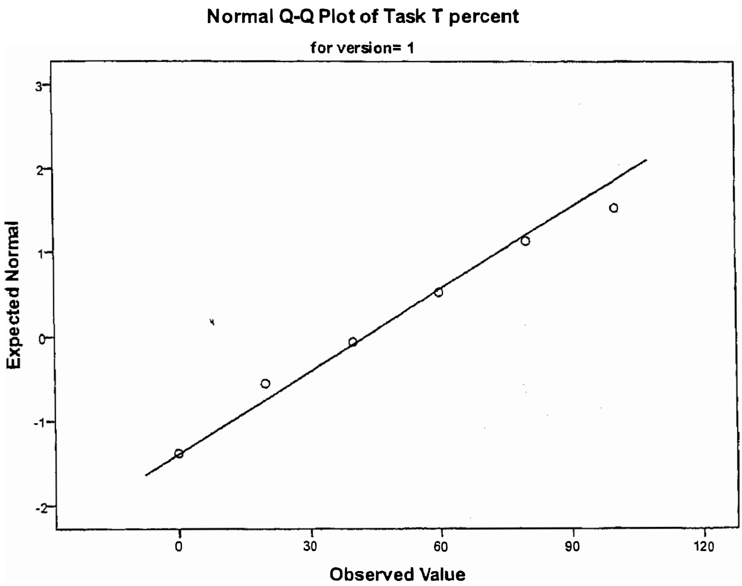


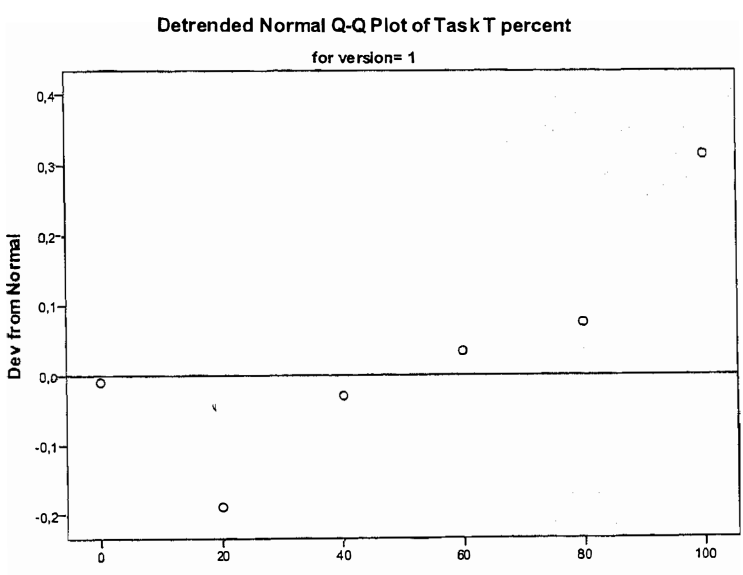




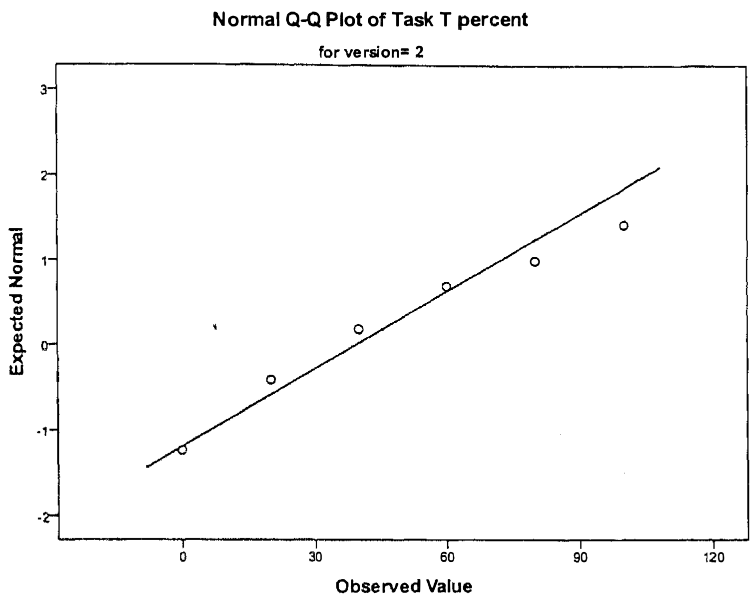
Task S percent
100,00
80,00
60,00
40,00
20,00
0,00
1
2
3
version

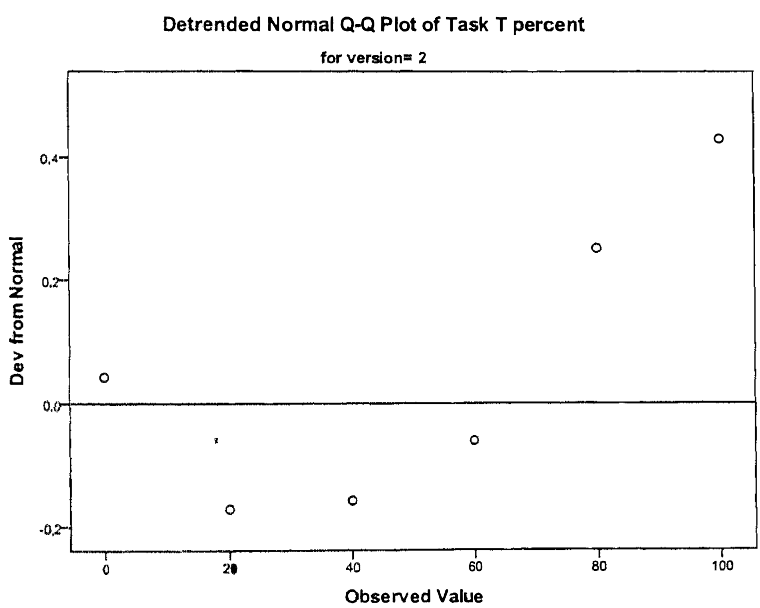
Version 1 Task T



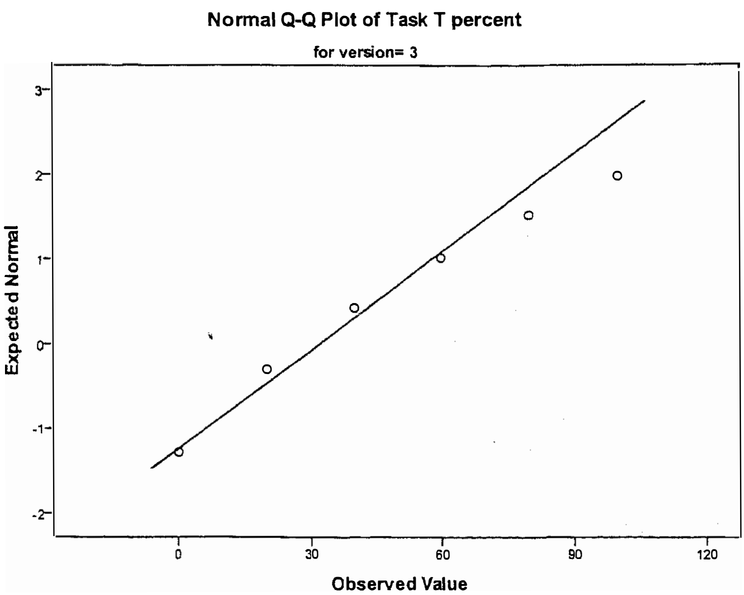


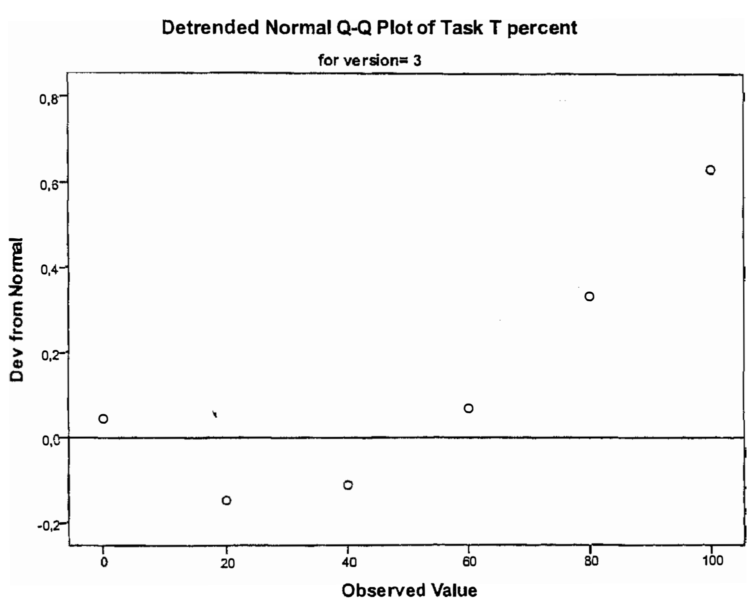
Version 2 Task T

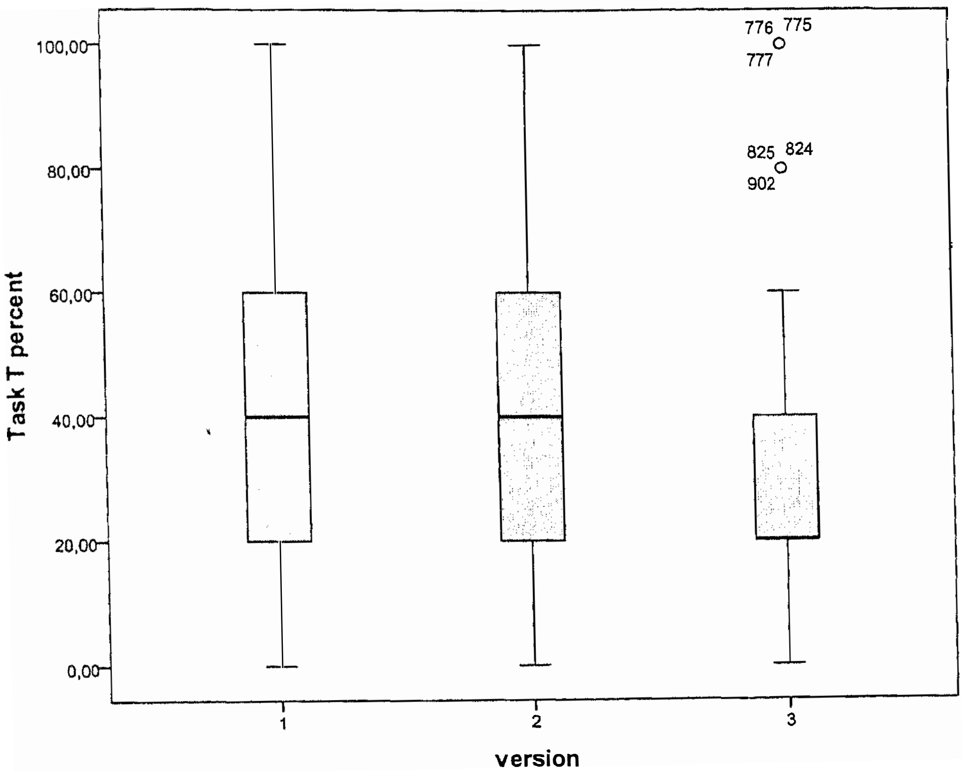




Version 3 Task T

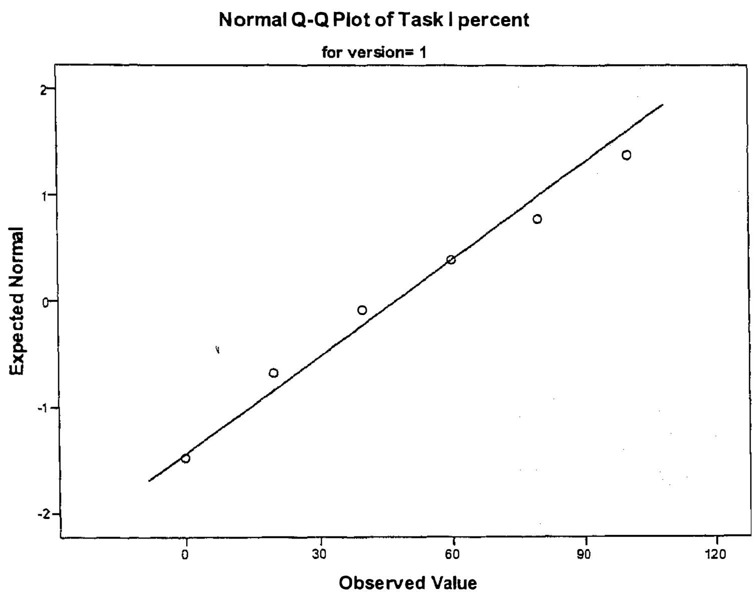


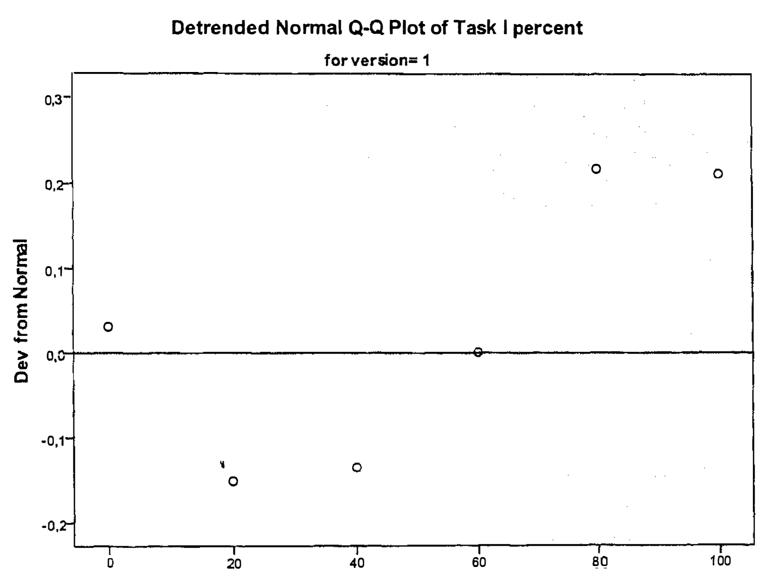




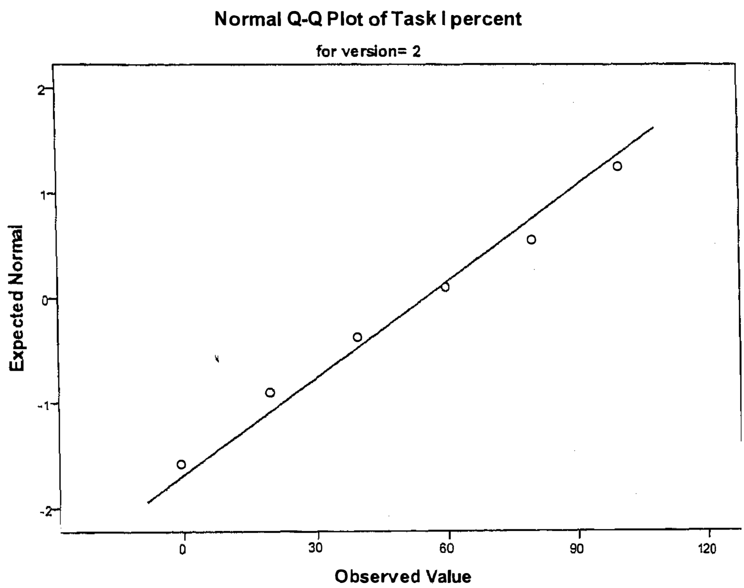
776
775
777
825
824
902
100,00
80,00
60,00
40,00
20,00
0,00
Task T percent
1
2
3
version

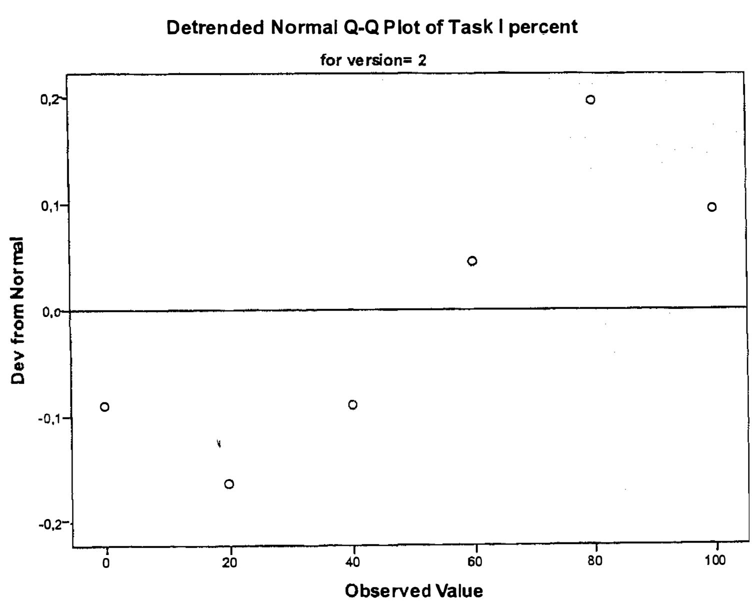
Version 1 Task I



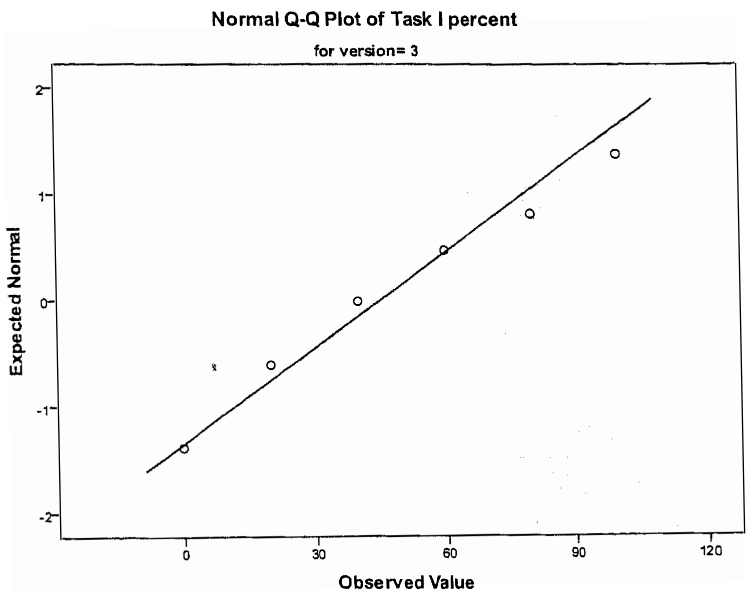


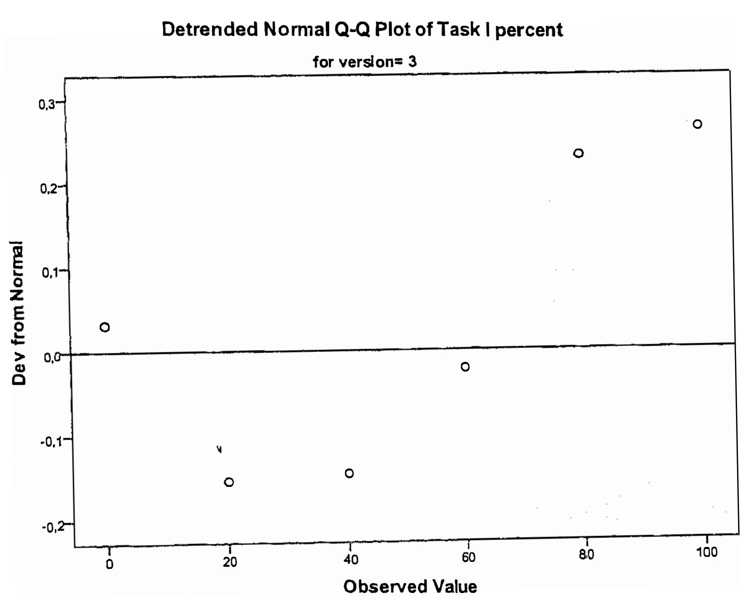
Version 2 Task I

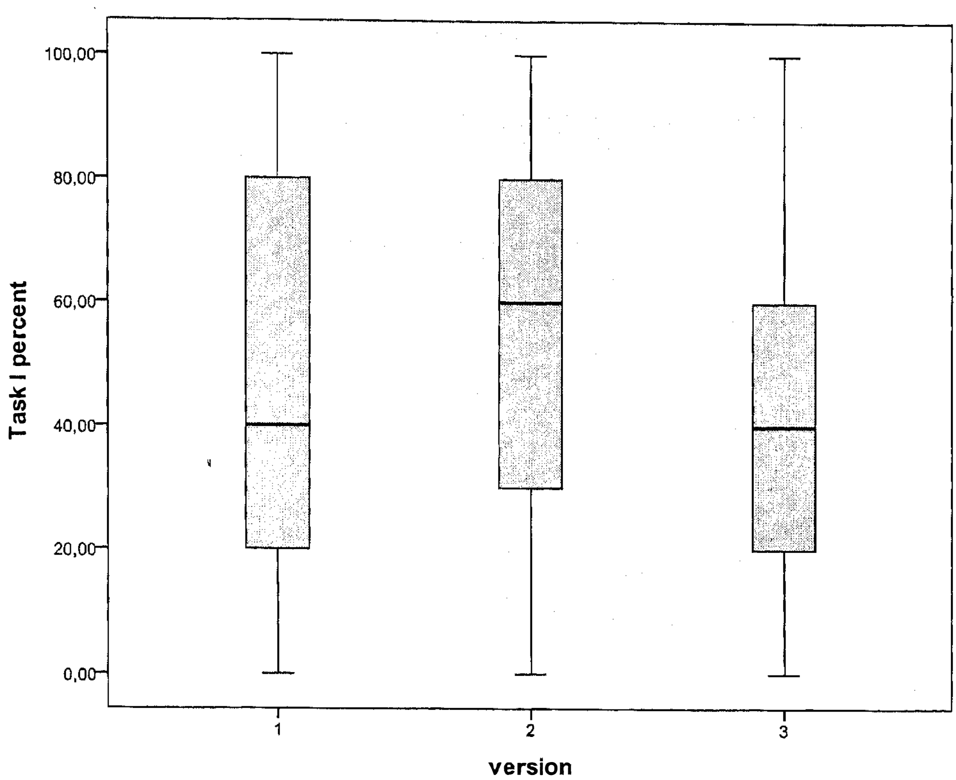




Version 3 Task I





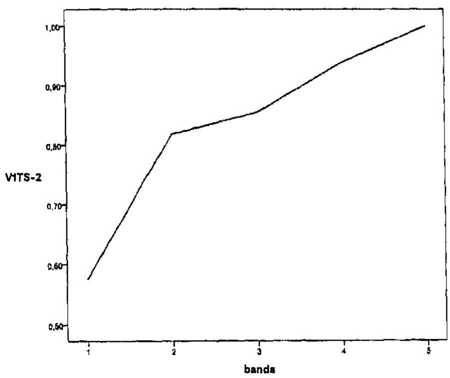

100,00
80,00
60,00
40,00
20,00
0,00
Task I percent
1
2
3
version

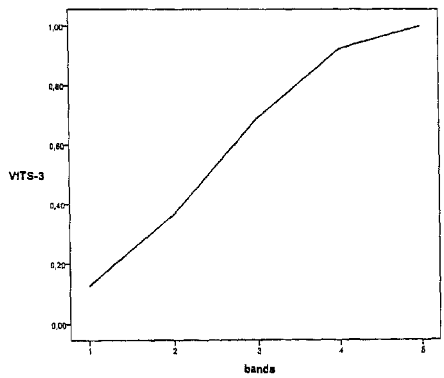
Appendix 4.3

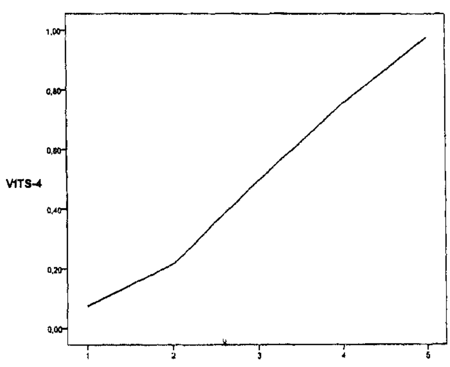
Item Discrimination Patterns and Graphs: Task S

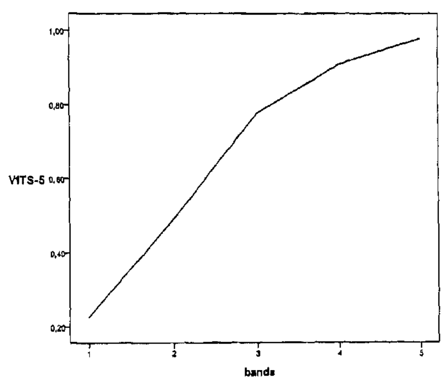
Item mean distributions to bands V1 TS

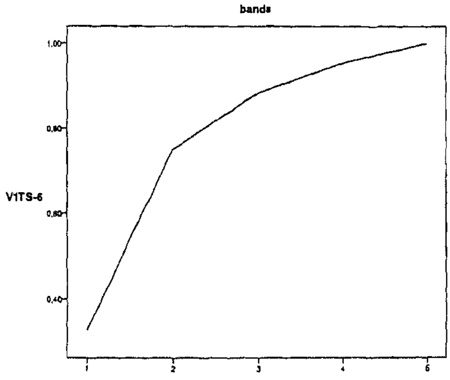
Band	V1TS-2	V1TS-3	V1TS-4	V1TS-5	V1TS-6	V1TS-7
1	0.58	0.13	0.08	0.23	0.33	0.63
2	0.82	0.36	0.22	0.49	0.75	0.78
3	0.85	0.69	0.50	0.78	0.88	0.97
4	0.94	0.92	0.75	0.91	0.95	0.98
5	1.00	1.00	0.98	0.98	1.00	1.00

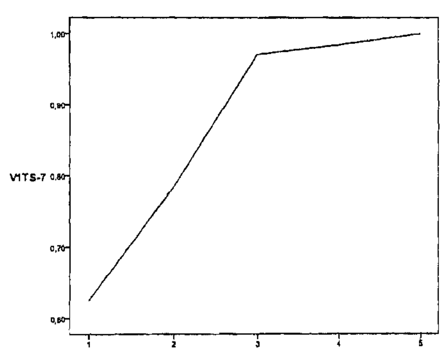






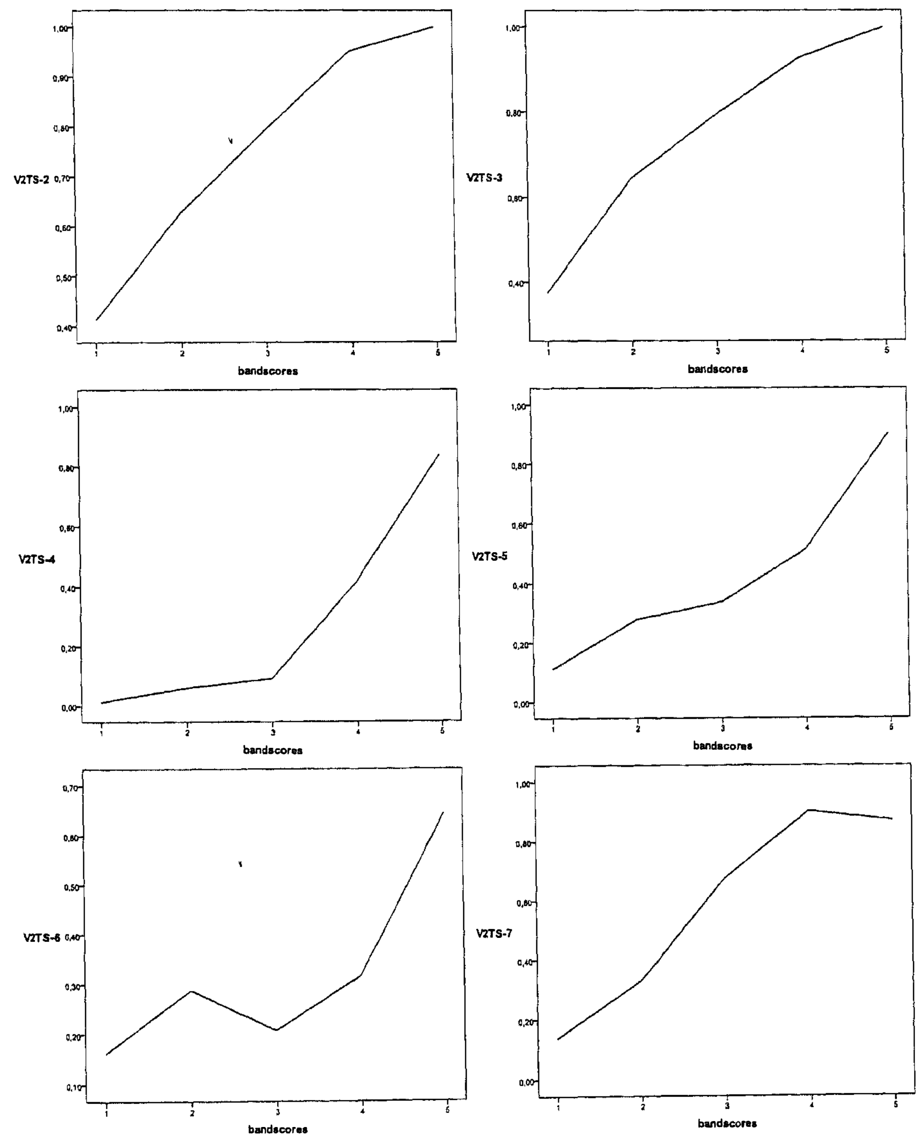






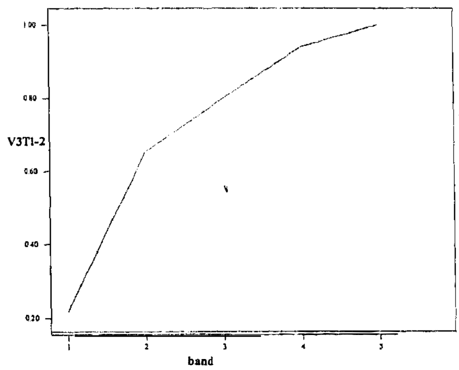
Band	V2TS-2	V2TS-3	V2TS-4	V2TS-5	V2TS-6	V2TS-7
1	0.41	0.38	0.01	0.14	0.16	0.14
2	0.63	0.64	0.06	0.28	0.29	0.33
3	0.79	0.79	0.09	0.34	0.21	0.68
4	0.95	0.93	0.41	0.51	0.32	0.90
5	1.00	1.00	0.84	0.90	0.65	0.87

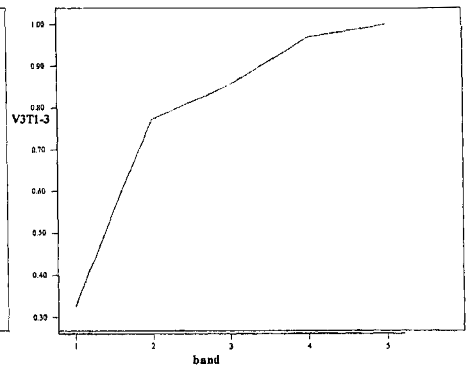
Item mean distributions to bands Version 2

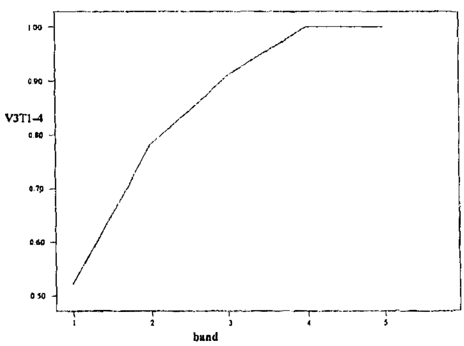


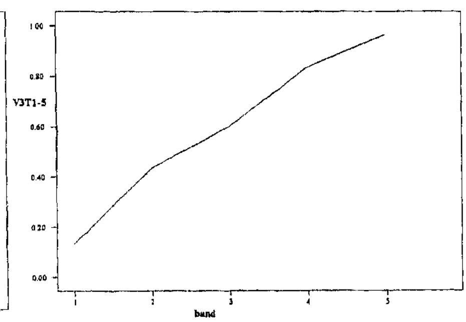
Item mean distributions to bands V3 TS

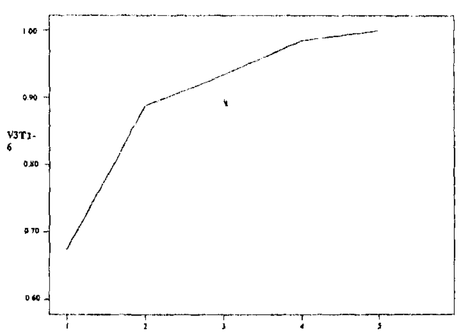
Band	V3T1-2	V3T1-3	V3T1-4	V3T1-5	V3T1-6	V3T1-7
1	0.22	0.33	0.52	0.13	0.67	0.04
2	0.65	0.77	0.78	0.43	0.89	0.33
3	0.80	0.86	0.91	0.60	0.93	0.63
4	0.94	0.97	1.00	0.83	0.98	0.89
5	1.00	1.00	1.00	0.96	1.00	0.96

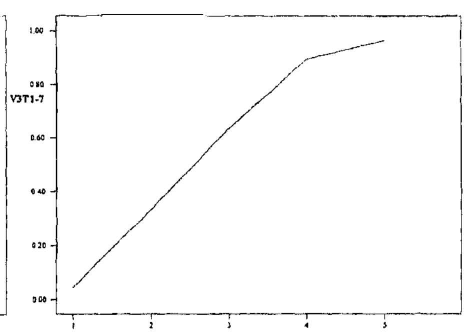










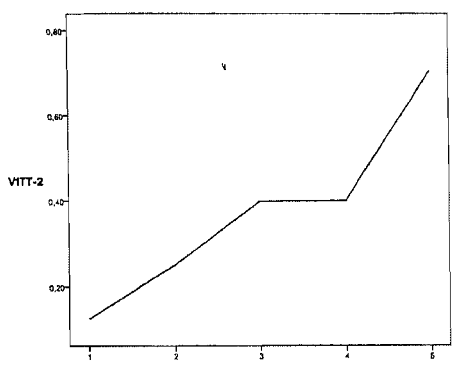


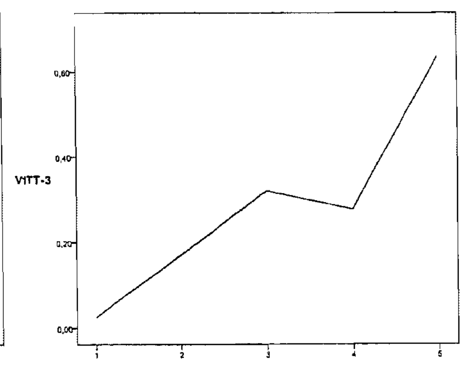
Appendix 4.4

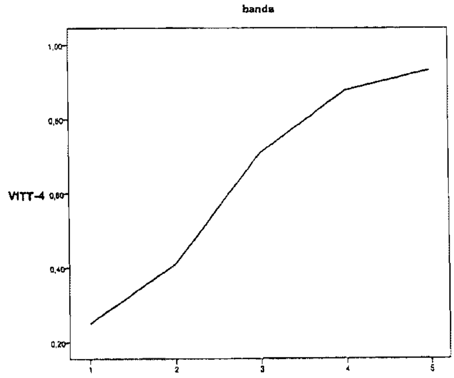
Item Discrimination Patterns and Graphs: Task T

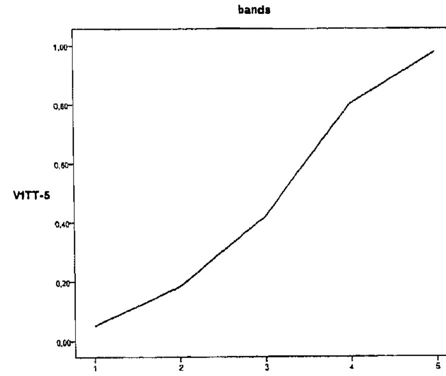
Item mean distributions to bands V1 TT

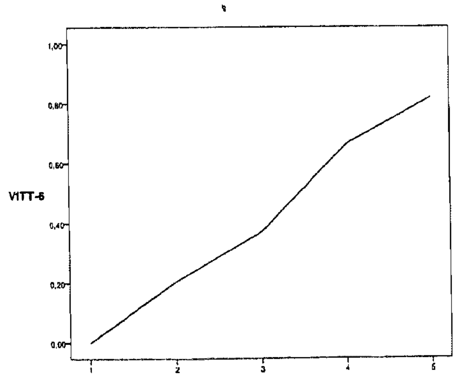
Band	**V1T2-2**	**V1T2-3**	**V1T2-4**	**V1T2-5**	**V1T2-6**
1	0.13	0.03	0.25	0.05	0.00
2	0.25	0.17	0.41	0.18	0.20
3	0.40	0.32	0.71	0.42	0.37
4	0.40	0.28	0.88	0.80	0.66
5	0.70	0.64	0.93	0.98	0.82





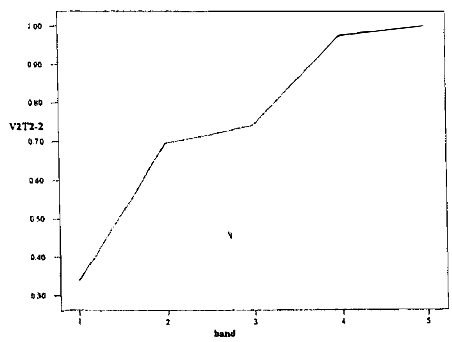


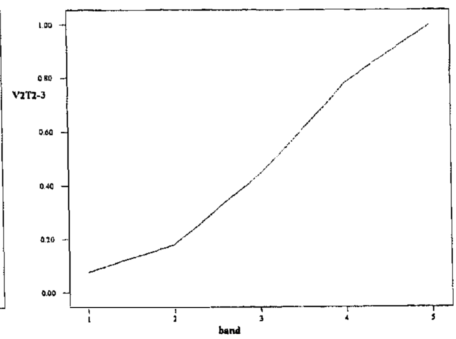


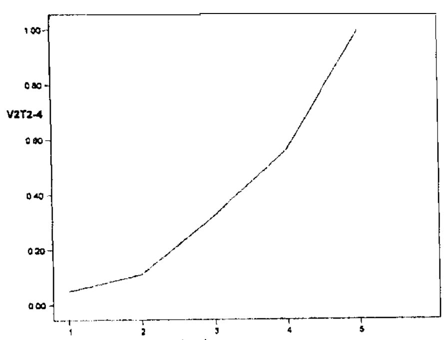


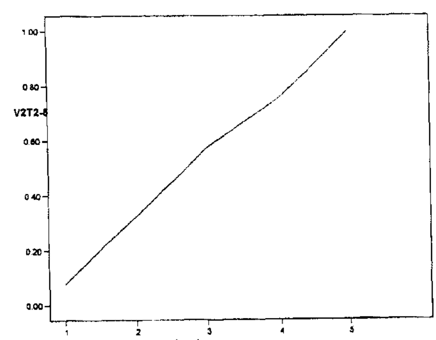
Band	V2T2-2	V2T2-3	V2T2-4	V2T2-5	V2T2-6
1	0.34	0.08	0.05	0.08	0.01
2	0.69	0.18	0.11	0.32	0.05
3	0.74	0.44	0.32	0.57	0.17
4	0.98	0.78	0.56	0.76	0.56
5	1.00	1.00	1.00	1.00	1.00

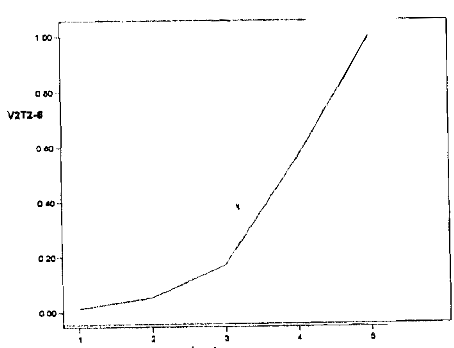
Item mean distributions to bands V2 TT





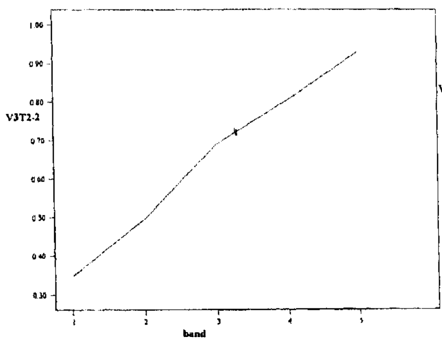


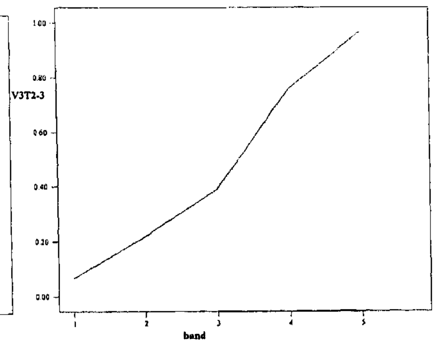


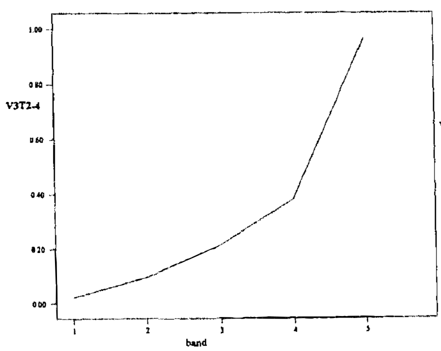


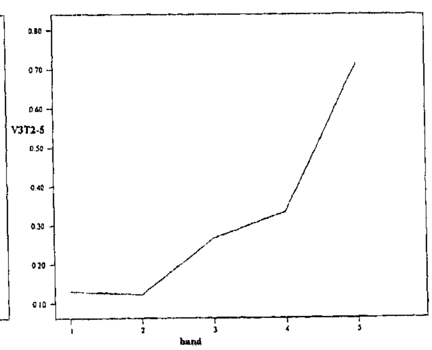
Item mean distributions to bands V3 TT

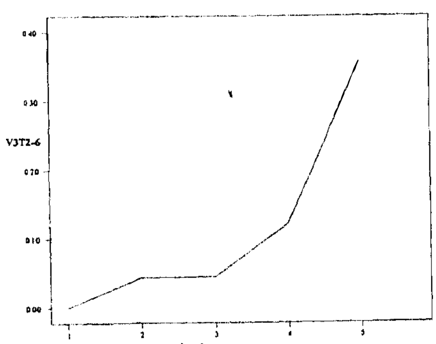
Band	V3T2-2	V3T2-3	V3T2-4	V3T2-5	V3T2-6
1	0.35	0.07	0.02	0.13	0.00
2	0.50	0.22	0.10	0.12	0.04
3	0.69	0.39	0.21	0.27	0.04
4	0.80	0.76	0.38	0.33	0.12
5	0.93	0.96	0.96	0.71	0.36









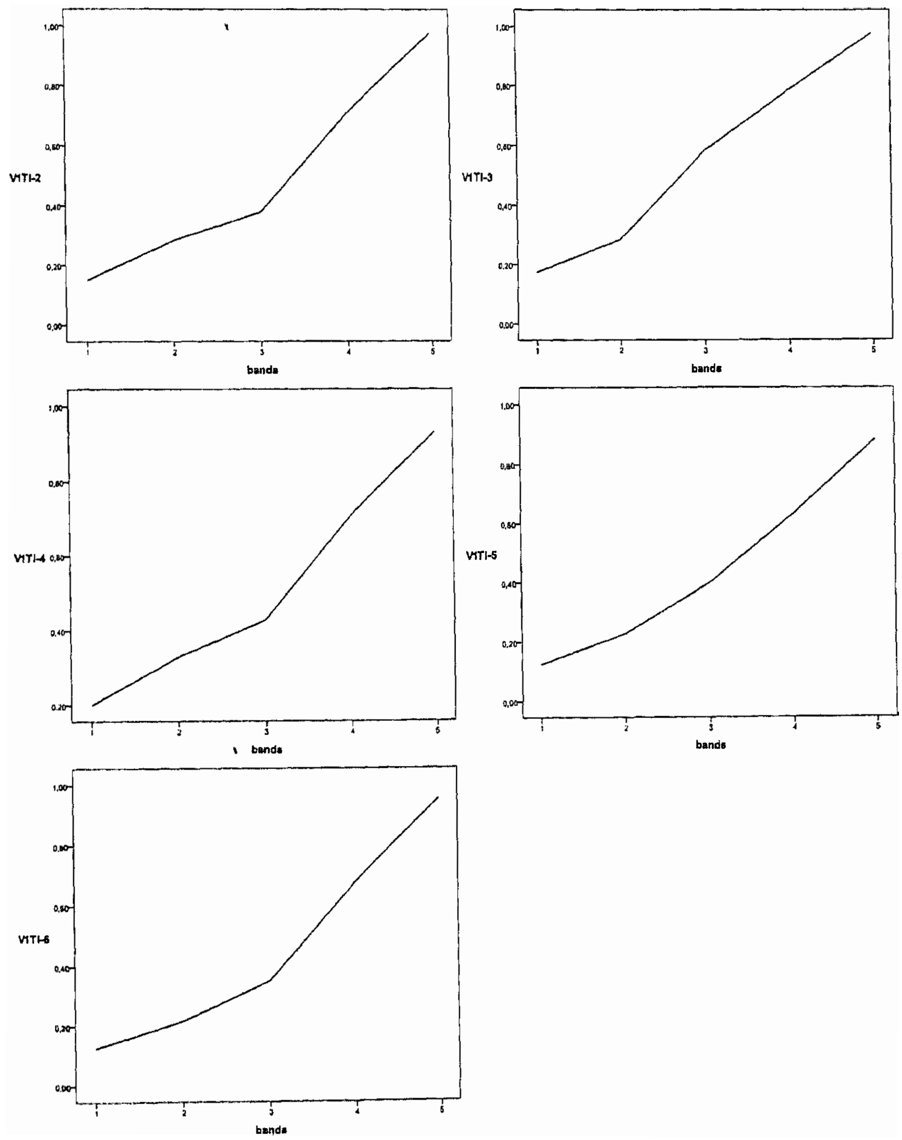


Appendix 4.5

Item Discrimination Patterns and Graphs: Task I

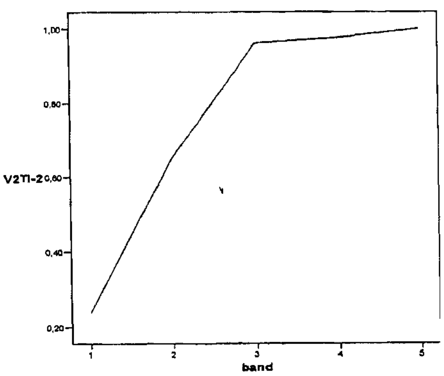
Item mean distributions to bands V1 TI

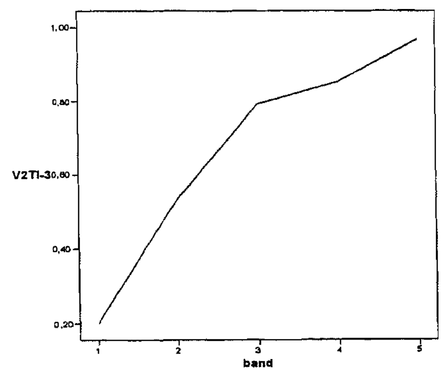
Band	**V1TI-2**	**V1TI-3**	**V1TI-4**	**V1TI-5**	**V1TI-6**
1	0.15	0.18	0.20	0.13	0.13
2	0.28	0.28	0.33	0.23	0.22
3	0.38	0.58	0.43	0.40	0.35
4	0.71	0.78	0.71	0.63	0.68
5	0.98	0.98	0.93	0.89	0.95

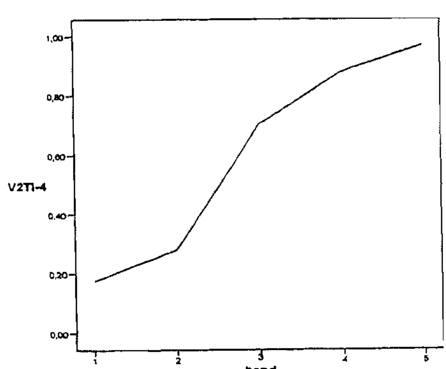


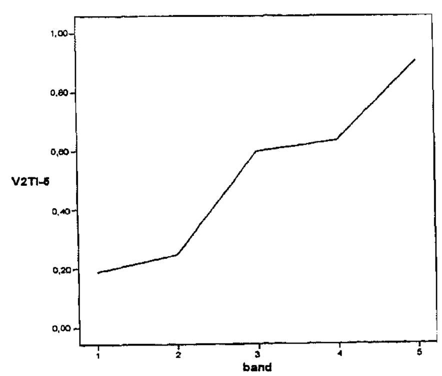
Item mean distributions to bands V2 TI

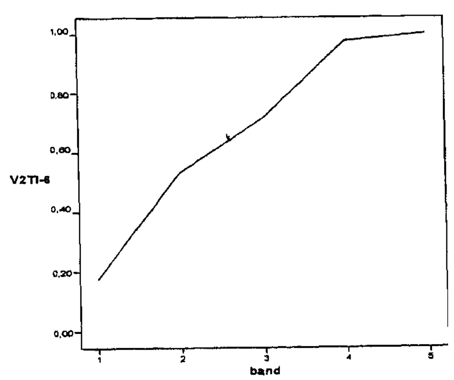
Band	**V2TI-2**	**V2TI-3**	**V2TI-4**	**V2TI-5**	**V2TI-6**
1	0.24	0.20	0.18	0.19	0.18
2	0.65	0.53	0.28	0.25	0.53
3	0.96	0.79	0.70	0.60	0.71
4	0.98	0.85	0.88	0.63	0.98
5	1.00	0.97	0.97	0.90	1.00





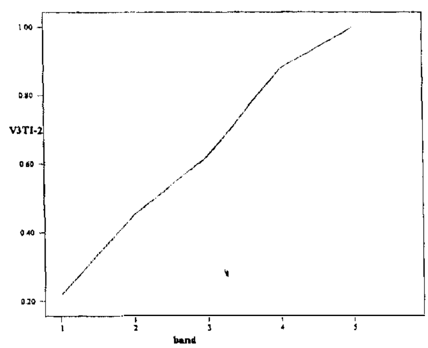


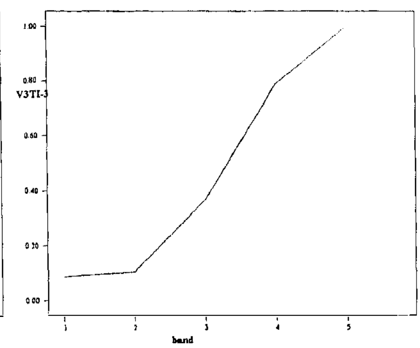


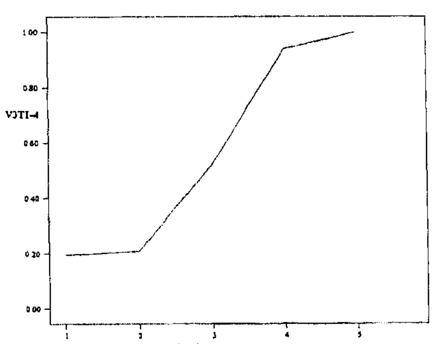


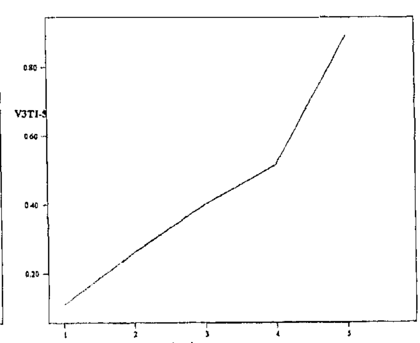
Item mean distributions to bands V3 TI

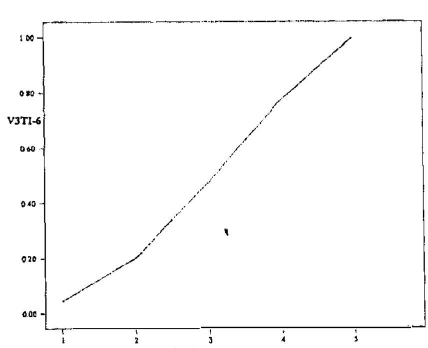
Band	**V3TI-2**	**V3TI-3**	**V3TI-4**	**V3TI-5**	**V3TI-6**
1	0.22	0.09	0.20	0.11	0.04
2	0.45	0.10	0.21	0.26	0.20
3	0.62	0.37	0.52	0.40	0.48
4	0.88	0.79	0.94	0.52	0.77
5	1.00	1.00	1.00	0.89	1.00











Appendix 4.6

Principal Axis Factoring Results by Version

Version 1

Descriptive Statistics Version 1

	Mean	Std. Deviation	Analysis N
TaskS_2	,85	,360	340
TaskS_3	,62	,485	340
TaskS_4	,49	,501	340
TaskS_5	,69	,464	340
TaskS_6	,81	,391	340
TaskS_7	,89	,316	340
TaskT_2	,37	,483	340
TaskT_3	,28	,449	340
TaskT_4	,64	,481	340
TaskT_5	,46	,499	340
TaskT_6	,40	,491	340
TaskI_2	,47	,500	340
TaskI_3	,55	,499	340
TaskI_4	,49	,501	340
TaskI_5	,43	,496	340
TaskI_6	,43	,496	340

KMO and Bartlett's Test Version 1

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,734
Bartlett's Test of Sphericity	Approx. Chi-Square	1309,309
	df	120
	Sig.	,000

Communalities Version 1

	Initial	Extraction
TaskS_2	,108	,074
TaskS_3	,454	,719
TaskS_4	,429	,501
TaskS_5	,276	,324
TaskS_6	,272	,445
TaskS_7	,167	,253
TaskT_2	,514	,759
TaskT_3	,510	,649
TaskT_4	,267	,266
TaskT_5	,535	,846
TaskT_6	,453	,511
TaskI_2	,346	,512
TaskI_3	,255	,304
TaskI_4	,269	,328
TaskI_5	,246	,280
TaskI_6	,238	,243

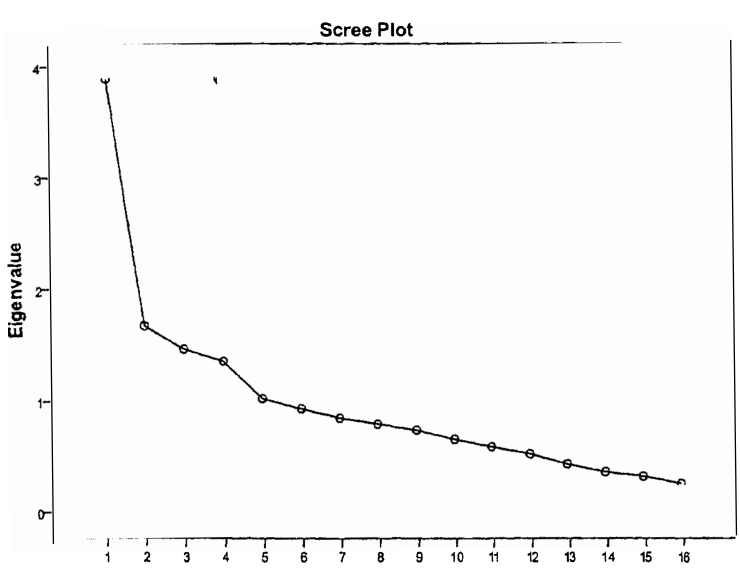
Extraction Method: Principal Axis Factoring.

Total Variance Explained Version 1

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Var.	Cum %	Total	% of Var.	Cum %	Total
1	3,887	24,292	24,292	3,361	21,007	21,007	2,264
2	1,681	10,505	34,797	1,366	8,537	29,544	2,272
3	1,476	9,224	44,021	,985	6,154	35,698	2,275
4	1,367	8,541	52,561	,867	5,416	41,114	1,515
5	1,038	6,487	59,049	,437	2,729	43,842	2,175
6	,945	5,907	64,956				
7	,861	5,381	70,337				
8	,805	5,032	75,369				
9	,749	4,682	80,051				
10	,668	4,172	84,223				
11	,598	3,739	87,962				
12	,532	3,325	91,287				
13	,440	2,751	94,038				
14	,367	2,295	96,332				
15	,327	2,046	98,379				
16	,259	1,621	100,000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.



Factor Matrix[a] Version 1

	Factor				
	1	2	3	4	5
TaskS_2	,243	-,004	,096	-,030	-,069
TaskS_3	,646	-,164	,344	-,212	-,334
TaskS_4	,582	-,132	,324	-,104	-,170
TaskS_5	,496	-,122	,175	-,108	,147
TaskS_6	,466	-,043	,227	-,268	,320
TaskS_7	,352	,022	,167	-,118	,293
TaskT_2	,224	,835	,086	,061	-,006
TaskT_3	,261	,753	,078	-,016	-,089
TaskT_4	,467	-,037	-,096	-,193	-,013
TaskT_5	,667	-,003	-,601	-,197	-,034
TaskT_6	,488	,005	-,489	-,162	-,086
TaskI_2	,476	-,127	,015	,514	-,072
TaskI_3	,454	-,004	-,119	,203	,208
TaskI_4	,398	-,108	-,015	,395	-,047
TaskI_5	,420	-,077	,060	,305	,029
TaskI_6	,413	,070	-,032	,229	,116

Extraction Method: Principal Axis Factoring.
a. Attempted to extract 5 factors. More than 25 iterations required. (Convergence=,002). Extraction was terminated.

Pattern Matrix[a] Version 1

	Factor				
	1	2	3	4	5
TaskS_2	,044	,017	,220	,051	,023
TaskS_3	-,021	,048	,855	,016	-,030
TaskS_4	,086	-,038	,621	,013	,098
TaskS_5	,080	,008	,197	-,066	,397
TaskS_6	-,102	-,030	,091	-,023	,665
TaskS_7	,013	-,065	-,023	,031	,529
TaskT_2	,030	-,039	-,038	,870	,023
TaskT_3	-,038	,036	,090	,803	-,037
TaskT_4	-,047	,344	,192	,001	,136
TaskT_5	,016	,928	-,021	-,009	-,015
TaskT_6	-,014	,751	,020	,004	-,094
TaskI_2	,743	-,047	,094	-,028	-,143
TaskI_3	,401	,143	-,178	,007	,239
TaskI_4	,585	,000	,062	-,033	-,101
TaskI_5	,495	-,049	,066	-,007	,051
TaskI_6	,397	,052	-,076	,105	,141

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Structure Matrix Version 1

	Factor				
	1	2	3	4	5
TaskS_2	,157	,128	,259	,074	,180
TaskS_3	,338	,321	,847	,058	,459
TaskS_4	,366	,255	,698	,066	,462
TaskS_5	,316	,275	,449	,026	,530
TaskS_6	,188	,251	,406	,085	,657
TaskS_7	,194	,171	,255	,117	,498
TaskT_2	,112	,099	,017	,870	,155
TaskT_3	,094	,153	,107	,801	,159
TaskT_4	,229	,451	,366	,080	,376
TaskT_5	,384	,920	,295	,126	,390
TaskT_6	,266	,710	,219	,096	,245
TaskI_2	,702	,224	,302	,032	,183
TaskI_3	,485	,355	,168	,110	,367
TaskI_4	,565	,213	,243	,021	,163
TaskI_5	,521	,199	,280	,057	,264
TaskI_6	,457	,268	,187	,182	,301

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix Version 1

Factor	1	2	3	4	5
1	1,000	,413	,409	,119	,403
2	,413	1,000	,342	,147	,444
3	,409	,342	1,000	,050	,554
4	,119	,147	,050	1,000	,181
5	,403	,444	,554	,181	1,000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Principal Axis Factoring Results by Version

Version 2

Descriptive Statistics Version 2

	Mean	Std. Deviation	Analysis N
TaskS_2	,69	,465	347
TaskS_3	,68	,467	347
TaskS_5	,34	,473	347
TaskS_7	,48	,500	347
TaskT_2	,68	,466	347
TaskT_3	,36	,480	347
TaskT_4	,28	,448	347
TaskT_5	,43	,496	347
TaskT_6	,21	,410	347
TaskI_2	,69	,461	347
TaskI_3	,59	,492	347
TaskI_4	,48	,500	347
TaskI_5	,41	,493	347
TaskI_6	,59	,493	347

KMO and Bartlett's Test Version 2

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,833
Bartlett's Test of Sphericity	Approx. Chi-Square	1072,245
	df	91
	Sig.	,000

Communalities Version 2

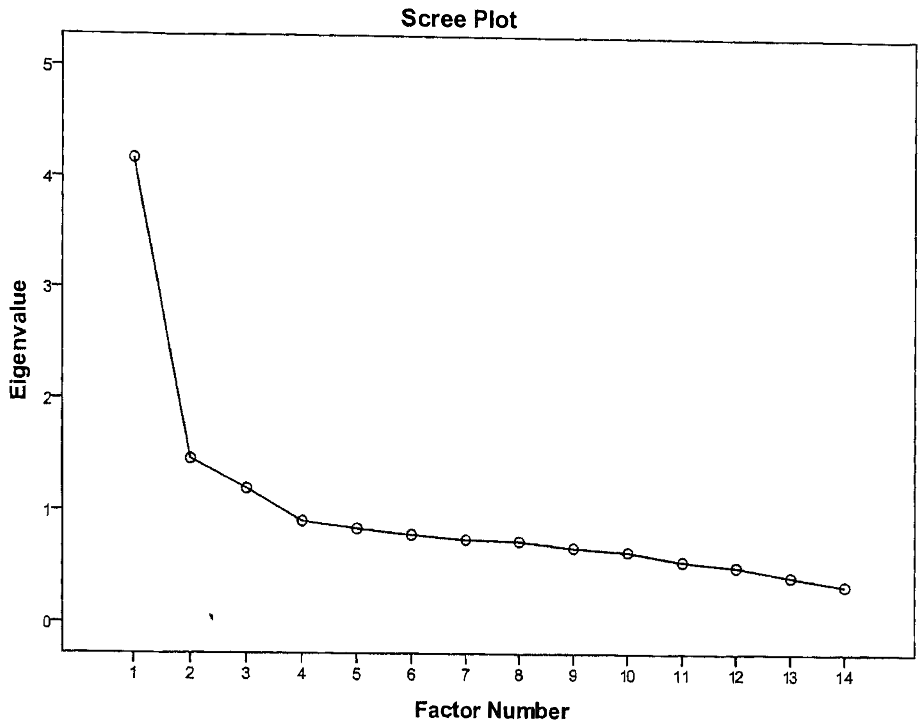
	Initial	Extraction
TaskS_2	,301	,477
TaskS_3	,332	,518
TaskS_5	,136	,152
TaskS_7	,218	,252
TaskT_2	,218	,219
TaskT_3	,445	,485
TaskT_4	,379	,461
TaskT_5	,307	,315
TaskT_6	,452	,599
TaskI_2	,327	,394
TaskI_3	,200	,241
TaskI_4	,295	,362
TaskI_5	,180	,219
TaskI_6	,290	,385

Extraction Method: Principal Axis Factoring.

Total Variance Explained Version 2

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Var	Cum%	Total	% of Var	Cum%	Total
1	4,169	29,780	29,780	3,556	25,402	25,402	2,978
2	1,467	10,475	40,256	,926	6,612	32,014	2,990
3	1,201	8,576	48,831	,599	4,278	36,293	1,660
4	,907	6,481	55,312				
5	,842	6,012	61,324				
6	,789	5,638	66,962				
7	,747	5,337	72,299				
8	,733	5,235	77,534				
9	,675	4,822	82,356				
10	,640	4,572	86,928				
11	,556	3,969	90,897				
12	,510	3,642	94,539				
13	,421	3,010	97,549				
14	,343	2,451	100,000				

Extraction Method: Principal Axis Factoring.
a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.



Factor Matrix[a] Version 2

	Factor		
	1	2	3
TaskS_2	,328	,567	-,220
TaskS_3	,375	,592	-,165
TaskS_5	,372	,082	,085
TaskS_7	,476	,154	,040
TaskT_2	,393	-,214	-,138
TaskT_3	,630	-,254	-,154
TaskT_4	,597	-,247	-,208
TaskT_5	,544	-,125	-,060
TaskT_6	,676	-,159	-,342
TaskI_2	,562	,014	,278
TaskI_3	,449	,004	,199
TaskI_4	,561	,032	,217
TaskI_5	,405	-,047	,229
TaskI_6	,535	,088	,303

Extraction Method: Principal Axis Factoring.
a. 3 factors extracted. 9 iterations required.

Pattern Matrix[a] Version 2

	Factor		
	1	2	3
TaskS_2	-,008	-,054	,713
TaskS_3	-,055	,042	,719
TaskS_5	,053	,302	,100
TaskS_7	,101	,304	,210
TaskT_2	,487	-,001	-,075
TaskT_3	,653	,095	-,068
TaskT_4	,684	,010	-,041
TaskT_5	,428	,183	-,003
TaskT_6	,795	-,109	,123
TaskI_2	,010	,632	-,029
TaskI_3	,037	,474	-,019
TaskI_4	,058	,555	,019
TaskI_5	,020	,485	-,091
TaskI_6	-,078	,657	,026

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Structure Matrix Version 2

	Factor		
	1	2	3
TaskS_2	,186	,236	,689
TaskS_3	,201	,303	,719
TaskS_5	,277	,377	,241
TaskS_7	,362	,455	,368
TaskT_2	,462	,279	,080
TaskT_3	,692	,484	,180
TaskT_4	,678	,431	,183
TaskT_5	,544	,455	,210
TaskT_6	,764	,449	,332
TaskI_2	,405	,627	,235
TaskI_3	,334	,490	,189
TaskI_4	,418	,600	,267
TaskI_5	,301	,460	,116
TaskI_6	,350	,618	,273

Factor Correlation Matrix

Factor	1	2	3
1	1,000	,639	,320
2	,639	1,000	,413
3	,320	,413	1,000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Principal Axis Factoring Results by Version

Version 3

Descriptive Statistics Version 3

	Mean	Std. Deviation	Analysis N
TaskS_2	,72	,452	345
TaskS_3	,79	,407	345
TaskS_4	,84	,367	345
TaskS_5	,56	,498	345
TaskS_6	,90	,302	345
TaskS_7	,53	,500	345
TaskT_2	,62	,486	345
TaskT_3	,41	,492	345
TaskT_4	,24	,428	345
TaskT_5	,25	,433	345
TaskT_6	,08	,269	345
TaskI_2	,59	,492	345
TaskI_3	,37	,485	345
TaskI_4	,49	,501	345
TaskI_5	,38	,485	345
TaskI_6	,43	,495	345

KMO and Bartlett's Test Version 3

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,802
Bartlett's Test of Sphericity	Approx. Chi-Square	935,904
	df	120
	Sig.	,000

Communalities Version 3

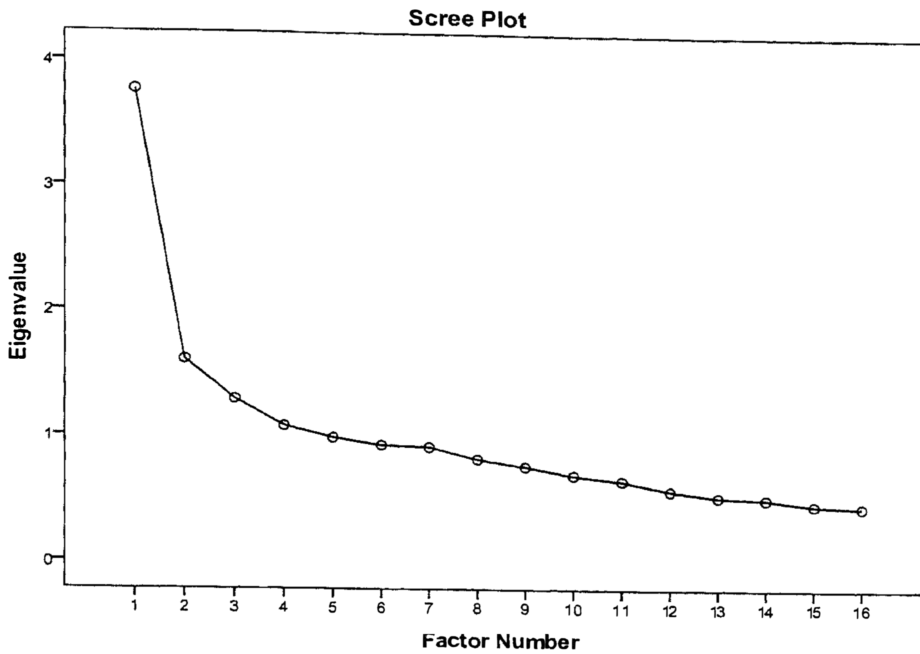
	Initial	Extraction
TaskS_2	,224	,260
TaskS_3	,233	,241
TaskS_4	,221	,298
TaskS_5	,219	,231
TaskS_6	,130	,146
TaskS_7	,370	,492
TaskT_2	,135	,138
TaskT_3	,293	,489
TaskT_4	,281	,384
TaskT_5	,192	,576
TaskT_6	,127	,153
TaskI_2	,219	,292
TaskI_3	,390	,541
TaskI_4	,332	,437
TaskI_5	,172	,190
TaskI_6	,303	,323

Extraction Method: Principal Axis Factoring.

Total Variance Explained Version 3

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Var	Cum %	Total	% of Var	Cum %	Total
1	3,762	23,512	23,512	3,134	19,586	19,586	2,296
2	1,602	10,009	33,521	,920	5,750	25,336	2,519
3	1,288	8,049	41,570	,715	4,466	29,802	2,142
4	1,077	6,732	48,303	,423	2,642	32,444	1,243
5	,985	6,154	54,457				
6	,927	5,793	60,250				
7	,913	5,707	65,957				
8	,819	5,119	71,076				
9	,762	4,759	75,836				
10	,695	4,341	80,177				
11	,652	4,074	84,251				
12	,575	3,591	87,842				
13	,526	3,289	91,130				
14	,510	3,190	94,320				
15	,462	2,888	97,208				
16	,447	2,792	100,000				

Extraction Method: Principal Axis Factoring.
a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.



Factor Matrix[a] Version 3

	Factor			
	1	2	3	4
TaskS_2	,431	,247	,088	-,076
TaskS_3	,403	,278	,013	-,033
TaskS_4	,348	,348	,119	,204
TaskS_5	,407	,241	,087	-,020
TaskS_6	,249	,284	,046	,041
TaskS_7	,582	,384	,074	,003
TaskT_2	,285	-,115	,040	-,205
TaskT_3	,548	-,094	,078	-,417
TaskT_4	,470	-,277	,278	-,096
TaskT_5	,309	-,321	,552	,269
TaskT_6	,273	-,239	,146	-,019
TaskI_2	,418	-,220	-,260	,024
TaskI_3	,641	-,162	-,274	,171
TaskI_4	,579	-,160	-,274	-,011
TaskI_5	,328	-,115	-,180	,192
TaskI_6	,546	-,049	-,123	,090

Extraction Method: Principal Axis Factoring.
a. Attempted to extract 4 factors. More than 25 iterations required. (Convergence=,005). Extraction was terminated.

Pattern Matrix[a] Version 3

	Factor			
	1	2	3	4
TaskS_2	,447	-,052	,164	-,014
TaskS_3	,453	,019	,079	-,071
TaskS_4	,583	,023	-,226	,095
TaskS_5	,442	-,021	,089	,014
TaskS_6	,420	-,034	-,049	-,033
TaskS_7	,663	,019	,068	-,025
TaskT_2	-,030	,010	,370	,024
TaskT_3	,071	-,058	,706	-,021
TaskT_4	-,027	,026	,377	,374
TaskT_5	,018	-,030	-,049	,779
TaskT_6	-,094	,092	,190	,257
TaskI_2	-,131	,550	,086	-,043
TaskI_3	,047	,737	-,055	,025
TaskI_4	-,013	,585	,153	-,075
TaskI_5	-,001	,499	-,157	,046
TaskI_6	,154	,446	,023	,045

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Structure Matrix

	Factor			
	1	2	3	4
TaskS_2	,494	,259	,332	,124
TaskS_3	,482	,265	,273	,059
TaskS_4	,512	,198	,087	,151
TaskS_5	,475	,250	,282	,134
TaskS_6	,374	,130	,111	,032
TaskS_7	,698	,374	,373	,148
TaskT_2	,148	,224	,370	,147
TaskT_3	,360	,393	,696	,219
TaskT_4	,239	,349	,509	,505
TaskT_5	,150	,177	,207	,757
TaskT_6	,093	,236	,291	,329
TaskI_2	,165	,526	,342	,119
TaskI_3	,384	,734	,417	,231
TaskI_4	,324	,648	,473	,145
TaskI_5	,179	,418	,158	,138
TaskI_6	,391	,548	,377	,217

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix

Factor	1	2	3	4
1	1,000	,484	,456	,217
2	,484	1,000	,600	,293
3	,456	,600	1,000	,342
4	,217	,293	,342	1,000

Extraction Method: Principal Axis Factoring.

Rotation Method: Promax with Kaiser Normalization.

Appendix 4.7

Principal Component Analysis: Version 1

Descriptive Statistics Version 1

	Mean	Std. Deviation	Analysis N
Task1_2	,85	,360	340
Task1_3	,62	,485	340
Task1_4	,49	,501	340
Task1_5	,69	,464	340
Task1_6	,81	,391	340
Task1_7	,89	,316	340
Task2_2	,37	,483	340
Task2_3	,28	,449	340
Task2_4	,64	,481	340
Task2_5	,46	,499	340
Task2_6	,40	,491	340
Task3_2	,47	,500	340
Task3_3	,55	,499	340
Task3_4	,49	,501	340
Task3_5	,43	,496	340
Task3_6	,43	,496	340

Correlation Matrix[a]

a. Determinant = ,020

KMO and Bartlett's Test Version 1

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,734
Bartlett's Test of Sphericity	Approx. Chi-Square	1309,309
	df	120
	Sig.	,000

Communalities Version 1

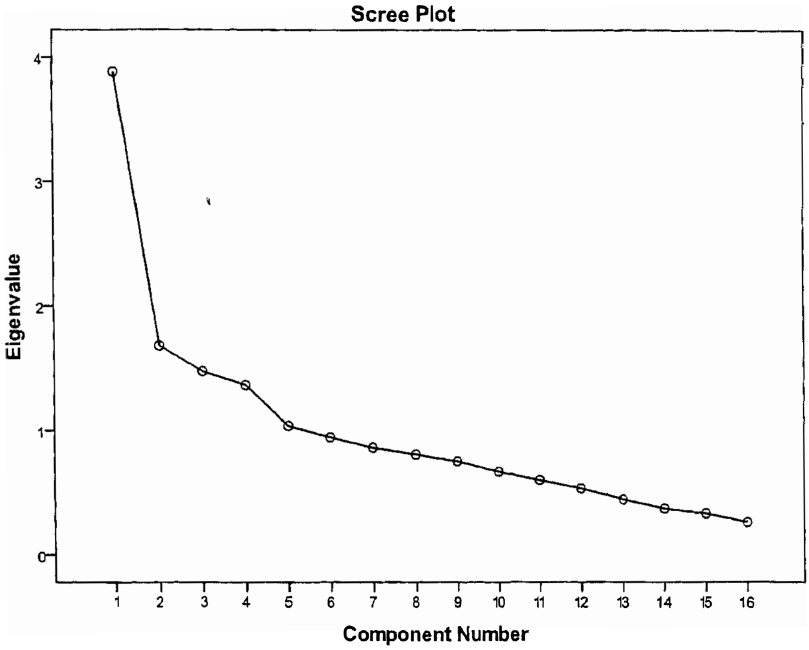
	Initial	Extraction
Task1_2	1,000	,409
Task1_3	1,000	,683
Task1_4	1,000	,589
Task1_5	1,000	,478
Task1_6	1,000	,556
Task1_7	1,000	,578
Task2_2	1,000	,843
Task2_3	1,000	,831
Task2_4	1,000	,457
Task2_5	1,000	,790
Task2_6	1,000	,743
Task3_2	1,000	,619
Task3_3	1,000	,498
Task3_4	1,000	,502
Task3_5	1,000	,440
Task3_6	1,000	,433

Extraction Method: Principal Component Analysis.

Total Variance Explained Version 1

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var	Cum%	Total	% of Var	Cum%	Total	% of Var	Cum%
1	3,887	24,292	24,292	3,887	24,292	24,292	2,210	13,810	13,810
2	1,681	10,505	34,797	1,681	10,505	34,797	1,998	12,489	26,299
3	1,476	9,224	44,021	1,476	9,224	44,021	1,927	12,043	38,343
4	1,367	8,541	52,561	1,367	8,541	52,561	1,729	10,808	49,151
5	1,038	6,487	59,049	1,038	6,487	59,049	1,584	9,898	59,049
6	,945	5,907	64,956						
7	,861	5,381	70,337						
8	,805	5,032	75,369						
9	,749	4,682	80,051						
10	,668	4,172	84,223						
11	,598	3,739	87,962						
12	,532	3,325	91,287						
13	,440	2,751	94,038						
14	,367	2,295	96,332						
15	,327	2,046	98,379						
16	,259	1,621	100,000						

Extraction Method: Principal Component Analysis.



Component Matrix[a] Version 1

	Component				
	1	2	3	4	5
Task1_2	,295	,005	-,138	,142	,531
Task1_3	,647	-,160	-,325	,154	,332
Task1_4	,621	-,162	-,284	,246	,186
Task1_5	,560	-,166	-,294	,134	-,180
Task1_6	,509	-,035	-,467	,100	-,261
Task1_7	,409	,049	-,324	,164	-,526
Task2_2	,214	,874	,039	,175	,027
Task2_3	,259	,855	-,031	,122	,130
Task2_4	,529	-,011	-,200	-,354	,107
Task2_5	,635	,048	,082	-,613	,050
Task2_6	,507	,060	,099	-,685	,050
Task3_2	,513	-,186	,481	,262	,146
Task3_3	,519	,016	,332	-,056	-,339
Task3_4	,451	-,173	,470	,199	,088
Task3_5	,485	-,128	,339	,270	-,036
Task3_6	,481	,120	,355	,086	-,233

Extraction Method: Principal Component Analysis.
a. 5 components extracted.

Rotated Component Matrix[a] Version 1

	Component				
	1	2	3	4	5
Task1_2	,056	,056	-,060	,106	,623
Task1_3	,164	,194	,360	-,010	,699
Task1_4	,225	,093	,432	-,010	,585
Task1_5	,183	,135	,601	-,064	,246
Task1_6	,003	,136	,707	,040	,193
Task1_7	,086	,022	,741	,103	-,102
Task2_2	,063	,010	,044	,915	,005
Task2_3	,010	,081	,034	,900	,117
Task2_4	,020	,569	,240	,020	,272
Task2_5	,197	,854	,103	,055	,092
Task2_6	,114	,853	,012	,033	,013
Task3_2	,743	,046	-,033	-,034	,251
Task3_3	,545	,297	,261	,078	-,194
Task3_4	,683	,065	-,038	-,046	,165
Task3_5	,633	,010	,141	,005	,139
Task3_6	,562	,169	,188	,203	-,102

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Component Transformation Matrix Version 1

Component	1	2	3	4	5
1	,556	,501	,504	,164	,399
2	-,161	,057	-,074	,970	-,155
3	,739	,031	-,575	,021	-,350
4	,332	-,861	,223	,165	,269
5	-,097	,065	-,600	,060	,789

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Component Score Coefficient Matrix Version 1

	Component				
	1	2	3	4	5
Task1_2	-,043	-,021	-,192	,061	,495
Task1_3	-,049	-,005	,051	-,032	,441
Task1_4	,005	-,075	,131	-,031	,336
Task1_5	-,002	-,035	,320	-,071	,032
Task1_6	-,109	-,024	,417	-,009	-,013
Task1_7	-,019	-,089	,508	,030	-,253
Task2_2	,006	-,050	-,012	,537	-,013
Task2_3	-,043	-,007	-,047	,527	,078
Task2_4	-,120	,293	,028	-,023	,114
Task2_5	-,027	,474	-,080	-,015	-,041
Task2_6	-,055	,504	-,116	-,023	-,074
Task3_2	,382	-,086	-,155	-,039	,118
Task3_3	,257	,088	,125	,010	-,296
Task3_4	,356	-,058	-,135	-,045	,057
Task3_5	,320	-,107	,001	-,018	,007
Task3_6	,278	,005	,067	,092	-,206

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

Version 1: Four Components

Component Matrix[a] Version 1 – 4 components

	Component			
	1	2	3	4
Task1_2	,295	,005	-,138	,142
Task1_3	,647	-,160	-,325	,154
Task1_4	,621	-,162	-,284	,246
Task1_5	,560	-,166	-,294	,134
Task1_6	,509	-,035	-,467	,100
Task1_7	,409	,049	-,324	,164
Task2_2	,214	,874	,039	,175
Task2_3	,259	,855	-,031	,122
Task2_4	,529	-,011	-,200	-,354
Task2_5	,635	,048	,082	-,613
Task2_6	,507	,060	,099	-,685
Task3_2	,513	-,186	,481	,262
Task3_3	,519	,016	,332	-,056
Task3_4	,451	-,173	,470	,199
Task3_5	,485	-,128	,339	,270
Task3_6	,481	,120	,355	,086

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Component Transformation Matrix Version 1 – 4 components

Component	1	2	3	4
1	,660	,551	,485	,160
2	-,149	-,161	,064	,973
3	-,660	,750	,040	,021
4	,326	,329	-,871	,162

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Component Score Coefficient Matrix Version 1 – 4 components

	Component			
	1	2	3	4
Task1_2	,146	,005	-,058	,030
Task1_3	,306	-,021	-,032	-,052
Task1_4	,306	,019	-,093	-,043
Task1_5	,273	-,022	-,030	-,061
Task1_6	,322	-,137	-,014	,006
Task1_7	,249	-,072	-,060	,060
Task2_2	-,017	,008	-,050	,536
Task2_3	,010	-,032	-,013	,520
Task2_4	,096	-,111	,286	-,029
Task2_5	-,079	-,021	,474	-,017
Task2_6	-,127	-,048	,505	-,024
Task3_2	-,049	,398	-,097	-,049
Task3_3	-,075	,227	,110	,029
Task3_4	-,071	,367	-,065	-,052
Task3_5	,007	,318	-,108	-,017
Task3_6	-,067	,258	,019	,105

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

Appendix 4.8

Principal Component Analysis: Version 2

Descriptive Statistics Version 2

	Mean	Std. Deviation	Analysis N
Task1_2	,69	,465	347
Task1_3	,68	,467	347
Task1_5	,34	,473	347
Task1_7	,48	,500	347
Task2_2	,68	,466	347
Task2_3	,36	,480	347
Task2_4	,28	,448	347
Task2_5	,43	,496	347
Task2_6	,21	,410	347
Task3_2	,69	,461	347
Task3_3	,59	,492	347
Task3_4	,48	,500	347
Task3_5	,41	,493	347
Task3_6	,59	,493	347

KMO and Bartlett's Test Version 2

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,833
Bartlett's Test of Sphericity	Approx. Chi-Square	1072,245
	df	91
	Sig.	,000

Communalities Version 2

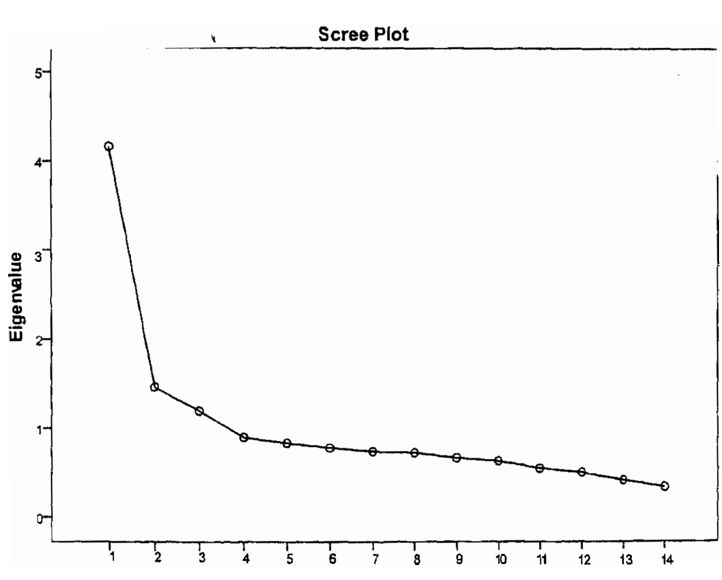
	Initial	Extraction
Task1_2	1,000	,709
Task1_3	1,000	,699
Task1_5	1,000	,224
Task1_7	1,000	,350
Task2_2	1,000	,434
Task2_3	1,000	,583
Task2_4	1,000	,574
Task2_5	1,000	,412
Task2_6	1,000	,641
Task3_2	1,000	,469
Task3_3	1,000	,373
Task3_4	1,000	,455
Task3_5	1,000	,413
Task3_6	1,000	,500

Extraction Method: Principal Component Analysis.

Total Variance Explained Version 2

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var	Cum%	Total	% of Var	Cum%	Total	% of Var	Cum%
1	4,169	29,780	29,780	4,169	29,780	29,780	2,563	18,309	18,309
2	1,467	10,475	40,256	1,467	10,475	40,256	2,558	18,274	36,583
3	1,201	8,576	48,831	1,201	8,576	48,831	1,715	12,248	48,831
4	,907	6,481	55,312						
5	,842	6,012	61,324						
6	,789	5,638	66,962						
7	,747	5,337	72,299						
8	,733	5,235	77,534						
9	,675	4,822	82,356						
10	,640	4,572	86,928						
11	,556	3,969	90,897						
12	,510	3,642	94,539						
13	,421	3,010	97,549						
14	,343	2,451	100,000						

Extraction Method: Principal Component Analysis.



Component Matrix[a] Version 2

	Component		
	1	2	3
Task1_2	,343	,680	,358
Task1_3	,390	,677	,299
Task1_5	,431	,162	-,108
Task1_7	,536	,250	-,003
Task2_2	,444	-,380	,304
Task2_3	,663	-,320	,203
Task2_4	,631	-,324	,268
Task2_5	,600	-,195	,137
Task2_6	,686	-,192	,364
Task3_2	,610	,031	-,310
Task3_3	,509	,025	-,337
Task3_4	,614	,064	-,272
Task3_5	,463	-,053	-,443
Task3_6	,582	,149	-,373

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

Rotated Component Matrix[a] Version 2

	Component		
	1	2	3
Task1_2	,049	,047	,839
Task1_3	,122	,049	,826
Task1_5	,389	,139	,231
Task1_7	,393	,215	,386
Task2_2	,037	,657	-,036
Task2_3	,265	,715	,043
Task2_4	,195	,730	,057
Task2_5	,299	,561	,088
Task2_6	,178	,746	,228
Task3_2	,642	,220	,093
Task3_3	,593	,143	,042
Task3_4	,621	,225	,138
Task3_5	,630	,098	-,085
Task3_6	,682	,105	,154

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Component Transformation Matrix Version 2

Component	1	2	3
1	,677	,657	,333
2	,113	-,539	,835
3	-,728	,527	,439

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Appendix 4.9

Principal Component Analysis: Version 3

Descriptive Statistics Version 3

	Mean	Std. Deviation	Analysis N
Task1_2	,72	,452	345
Task1_3	,79	,407	345
Task1_4	,84	,367	345
Task1_5	,56	,498	345
Task1_6	,90	,302	345
Task1_7	,53	,500	345
Task2_2	,62	,486	345
Task2_3	,41	,492	345
Task2_4	,24	,428	345
Task2_5	,25	,433	345
Task2_6	,08	,269	345
Task3_2	,59	,492	345
Task3_3	,37	,485	345
Task3_4	,49	,501	345
Task3_5	,38	,485	345
Task3_6	,43	,495	345

KMO and Bartlett's Test Version 3

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,802
Bartlett's Test of Sphericity	Approx. Chi-Square	935,904
	df	120
	Sig.	,000

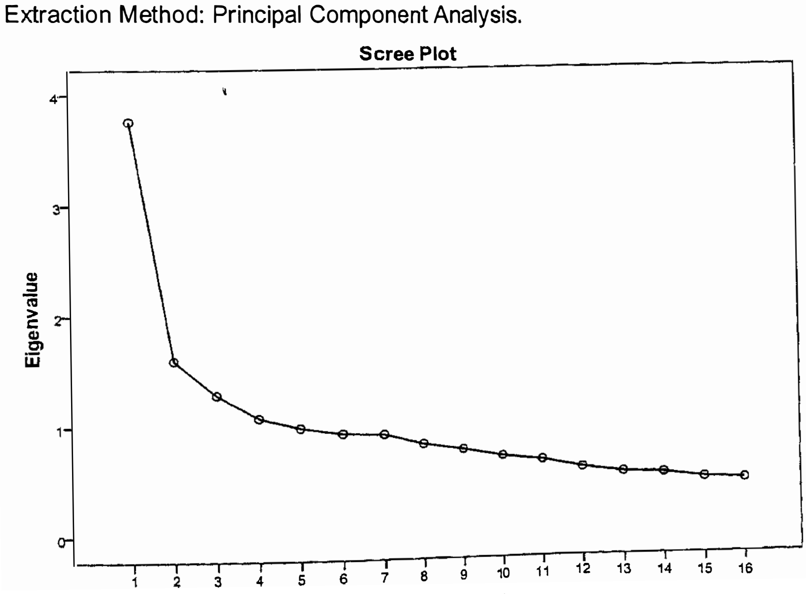
Communalities Version 3

	Initial	Extraction
Task1_2	1,000	,371
Task1_3	1,000	,363
Task1_4	1,000	,473
Task1_5	1,000	,395
Task1_6	1,000	,301
Task1_7	1,000	,554
Task2_2	1,000	,631
Task2_3	1,000	,487
Task2_4	1,000	,548
Task2_5	1,000	,630
Task2_6	1,000	,429
Task3_2	1,000	,489
Task3_3	1,000	,585
Task3_4	1,000	,536
Task3_5	1,000	,517
Task3_6	1,000	,420

Extraction Method: Principal Component Analysis.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var	Cum%	Total	% of Var	Cum%	Total	% of Var	Cum%
1	3,762	23,512	23,512	3,762	23,512	23,512	2,388	14,926	14,926
2	1,602	10,009	33,521	1,602	10,009	33,521	2,362	14,764	29,690
3	1,288	8,049	41,570	1,288	8,049	41,570	1,626	10,164	39,854
4	1,077	6,732	48,303	1,077	6,732	48,303	1,352	8,449	48,303
5	,985	6,154	54,457						
6	,927	5,793	60,250						
7	,913	5,707	65,957						
8	,819	5,119	71,076						
9	,762	4,759	75,836						
10	,695	4,341	80,177						
11	,652	4,074	84,251						
12	,575	3,591	87,842						
13	,526	3,289	91,130						
14	,510	3,190	94,320						
15	,462	2,888	97,208						
16	,447	2,792	100,000						

Extraction Method: Principal Component Analysis.



Component Matrix[a] Version 3

	Component			
	1	2	3	4
Task1_2	,493	,326	,113	-,095
Task1_3	,461	,371	-,038	-,107
Task1_4	,393	,492	,086	,263
Task1_5	,468	,336	,113	,224
Task1_6	,296	,448	,013	-,112
Task1_7	,617	,415	,017	,036
Task2_2	,337	-,191	,191	-,667
Task2_3	,578	-,119	,205	-,312
Task2_4	,510	-,314	,434	-,001
Task2_5	,315	-,268	,584	,343
Task2_6	,319	-,368	,351	,262
Task3_2	,469	-,353	-,379	-,041
Task3_3	,663	-,203	-,314	,076
Task3_4	,620	-,234	-,309	-,034
Task3_5	,381	-,214	-,384	,422
Task3_6	,607	-,080	-,191	-,095

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Rotated Component Matrix[a] Version 3

	Component			
	1	2	3	4
Task1_2	,560	,099	,079	,204
Task1_3	,557	,154	-,064	,156
Task1_4	,654	,022	,104	-,184
Task1_5	,578	,114	,200	-,091
Task1_6	,522	-,016	-,122	,111
Task1_7	,703	,218	,072	,082
Task2_2	,024	,102	,058	,785
Task2_3	,256	,257	,273	,530
Task2_4	,113	,183	,628	,328
Task2_5	,075	-,019	,790	-,008
Task2_6	-,034	,164	,633	,022
Task3_2	-,039	,683	,030	,140
Task3_3	,213	,718	,132	,084
Task3_4	,154	,688	,087	,176
Task3_5	,052	,624	,130	-,328
Task3_6	,276	,534	,075	,231

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Component Transformation Matrix Version 3

Component	1	2	3	4
1	,606	,643	,359	,300
2	,783	-,416	-,426	-,182
3	,115	-,634	,716	,268
4	,079	,110	,421	-,897

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

REFERENCES

Alderson, J.C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Alderson, J.C., Clapham, C. and Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Alderson, C., Figueras, N., Kuijper, H., Nold, G., Takala, S. and Tardieu, C. (2004). *The development of specification for item development and classification within the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Reading and Listening),* Final Report of the Dutch CEF construct project.

Alderson, C., Figueras, N., Kujper, H., Nold, G., Takala, S. and Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. . *Language Assessment Quarterly* 31ę3-30.

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

Bachman, L. F., Davidson, F., Ryan, K. and Choi, I. (1995). *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study.* Cambridge: Cambridge University Press.

Bachman, L. F., Kunnan, A., Vaniarjan, S., and Lynch, B. (1988). Task and ability analysis as a basis for examining content and construct comparability in two EFL batteries. *Language Testing.* 5, 128-159.

Bachman, L. F. and Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Barnett M.A. (1989). *More than meets the eye: Foreign language reading: Theory and practice.* New Jersey: Prentice Hall Regents.

Bereiter, C. and Scardamalia, M. (1987). *The psychology of written composition*, Hillside, NJ: Lawrence Erlbaum Associates.

Bhatia, V. K. (1997). Applied genre analysis and ESP. In Miller, T. (Ed.) *Functional approaches to written text: Classroom applications* 134–149. Washington, DC: USA.

Braten, I. and Samuelstuen, M. S. (2004). Does influence of reading purpose on reports of strategic text processing depend on students' topic knowledge?. *Journal of Educational Psychology* 96, 324-336.

Braten, I. and Stromso, H. I. (2003). A longitudinal think-aloud sudy of spontaneous strategic processing during the reading of multiple expository texts. *Reading and Writing: An Interdisciplinary Journal* 16, 195-218.

Braten, I and Stromso, H. I. (2006). Effects of personal epistemology on the understanding of multiple text. *Reading Psychology* 27, 457-484.

Braten, I., and Stromso, H. I. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes* 47, 1-31.

Braten, I., and Stromso, H.I. (in press). When law students read multiple documents about global warming: Examining the role of topic-specific beliefs about the nature of knowledge and knowing. *Instructional Science.*

Braten, I. and Stromso, H. I. and Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly* 44 (1), 6-28.

Braten, I., Stromso, H. I. and Samuelstuen, M. S. (2008). Are sophisticated students always better? The role of topic-specific personal epistemology in the understanding of multiple expository texts. *Contemporary Educational Psychology* 33, 814-840.

Britt, M. A. and Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction* 20 (4), 458-522.

Britt, M.tA. and Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology* 25, 313-339.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, New Jersey: Prentice Hall Regents.

Brown, J. D. (1997). An EFL readability index. *University of Hawaii Working Papers in English as a Second Language* 15, 85–1319.

Brown, J. D. and Hudson, T. (2002). *Criterion-referenced language testing.* Cambridge: Cambridge University Press.

Burstein, J. (2003). Thee-rater scoring engine: Automated essay scoring with natural language processing. In Shermis, M.D. and Burstein, J. (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. Hillsdale, NJ: Lawrence Erlbaum.

Carrell, P. L. (1984). The effects of rhetorical organisation on ESL readers. *TESOL Quarterly* 18, 441–469.

Carson, J. G. (2001). A task analysis of reading and writing in academic contexts. In D. Belcher and A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* 48-83. Ann Arbor: The Michigan University Press.

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

Carver, R. P. (1998). Predicting reading level in grades 1 to 6 from listening level and decoding level: Testing theory relevant to the simple view of reading. *Reading and Writing: An Interdisciplinary Journal* 10, 121-154.

Cerdan, R. (2006). Integration information processes from multiple documents. Unpublished PhD Thesis. *Departemento de Psicologia* Evolutiva y Educacion Valencia: Universitat de Valencia.

Chihara, T., Sakurai, T., and Oller, J. (1989). Background and culture as factors in EFL reading comprehension, *Language Testing* 6, 143–151.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman and R.R. Jordan (Eds.), *Reading in a second language* 5-12. MA: Newbury House Publishers.

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

Cohen, A. D. (1994b). Verbal reports on learning strategy. TESOL Quarterly 28, 673-697.

Cohen, A. D. (1998). *Strategies in learning and using a second language.* New York: Adisson Wesley Longman Limited.

Cohen, A. D. and Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks. TOEFL Monograph Series MS -33.* ETS.

Coltheart, M. (1981). The MRC psycholinguistic database quarterly. *Journal of Experimental Psychology* 33A, 497-505.

Council of Europe (2001). A common European framework of reference for languages: Learning, teaching and assessment. Retrieved on 10 May 2004 from: http://www.culture2.coe.int/portfolio/documents intro/common framework.html.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34, 213–238.

Crossley, S. A., Greenfield, J., and McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly* 42, 475–493.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M. And McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal* 91, 15–30.

Crossley, S. A. and McNamara, D. S. (2008). Assessing L2 reading texts at the intermediate level: An approximate replication of Crossley, Louwerse, McCarthy and McNamara (2007), *Language Teaching* 41, 409–429.

Cumming, A., Kantor, R., Powers, D., Santos, T. and Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper. TOEFL® Monograph* No. MS-18. ETS.

Edexcel (2004). *Pearson Test of English*. Retreived on 1 May 2006 from http://www.londontests.com.ar/lte.html

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P. and Schedl, M. (2000). *TOEFL 2000 Reading Framework: A Working Paper. TOEFL Monograph Series* MS-17. ETS.

Ehrlich, M. F. (1991). The processing of cohesion devices in text comprehension. *Psychological Research* 53, 169–174.

Fortus, R., Coriat, R., and Fund, S. (1998). Prediction of item difficulty in the English subtest of Israel's inter-university psychometric entrance test. In Kunnan, A. J. (Ed.), *Validation in language assessment: Selected papers from the 17th Language*

Research Colloquium, Long Beach 61–87. Mahwah, NJ: Lawrence Erlbaum.

Freedle, R. (1997). The relevance of multiple-choice reading test data in studying expository passage comprehension: The saga of a 15 year effort towards an experimental/correlational merger. *Discourse Processes* 23, 399-440.

Freedle, R., and Kostin, I. (1993). *The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main idea, inference, and supporting idea items.* TOEFL Research Reports No. RR-93–44. Princeton, NJ:. Educational Testing Service

Fulcher, G (2003). *Testing Second Language Speaking,* Harlow: Pearson.

Gernsbacher, M.A. (1990). *Language comprehension as structure building.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gil, L., Braten, I., Vidal-Abarca, E. and Strömso, H. I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology* 31,30-68.

Goh, S. T. (1990). The Effects of rhetorical organisation in expository prose on ESL readers in Singapore, *RELC Journal* 21, 1–13.

Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes* 23, 357-398.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. W. Shuart and D. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* 317-351. Information Age Publishing.

Goldman, S. R. and Bloome, D. M. (2004). Learning to construct and integrate. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications: Festshrift in honor of Lyle Bourne, Walter Kintsch and Thomas Landauer* 169-182. Washington, D. C.: American Psychological Association.

Goldman, S. R. and Rakestraw, J. A. (2000). Structural aspects of constructing meaning from text. In M L Kamil, P B Mosenthal, P D Pearson and R Barr (Eds.), *Handbook of reading research Vol. III* 311-335. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Grabe, W. (1999). Developments in reading research and their implications for computer-adaptive reading assessment. In M. Chalhoub-Deville (Ed.), *Issues in computer-adaptive testing of reading proficiency* 11-47. Cambridge: Cambridge University Press.

Grabe, W. (2000). Reading research and its implications for reading assessment. In A. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida.* Cambridge: Cambridge University Press.

Grabe, W. (2001). Reading-writing relations: Theoretical perspectives and instructional practices. In D. Belcher and A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* 15-47. Ann Arbor: The University of Michigan Press.

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics* 24, 44-69.

Grabe, W. and Stoller, F. L. (2002). Teaching and researching reading. Essex: Longman.

Graesser, A., McNamara, D., Louwerse, M., and Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers* 36, 193–202.

Green, A., Ünaldi, A. and Weir, C. (2008). *The cognitive processes of second language academic readers*. LLAS Pedagogic Research Fund Project Report. Retrieved on 1 September 2009 from http://www.llas.ac.uk/projects.

Green, A., Ünaldi, A. and Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. Language Testing 27, 191-214.

Green, R. and Weir, C. (1998): *SPSS guide for language testers.* Unpublished manuscript. Centre for Applied Linguistics, University of Reading.

Hale, G., Taylor, C., Bridgeman, B., Carson, J. , Kroll, B. and Kantor, R. (1996). A study of writing tasks assigned in academic degree programs. *TOEFL Research Reports* RR-95-44.

Hartmann (1995). Eight readers reading: The intertextual links of proficient readers reading multiple passages. *Reading Research Quarterly* 30, 520-561.

Hatch, E. and Lazarton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modelling.* Cary, North Carolina: SAS.

Henning, G. (1987). *A guide to language testing.* Cambridge, Massachusetts: New Jersey: Newbury House Publishers.

Horowitz, D. M. (1986). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly* 15 (3), 445-462.

Huhta, A., and Randell, E. (1996). Multiple-choice summary: A measure of text comprehension. In A. Cumming and R. Berwick (Eds.), *Validation in language testing* 94-110. Clevedon, England: Multilingual Matters.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Longman.

Jordan, R, R. (1998). *English for Academic Purposes: A guide and resource book for teachers.* Cambridge: Cambridge University Press.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review* 87 (4), 329-354.

Just, M. A. and Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Newton, Massachusetts: Allyn and Bacon, Inc.

Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99 (1), 122-149.

Khalifa, H. (1997). *A study in the construct validation of the reading module of an EAP proficiency test battery: Validation from a variety of perspectives.* Unpublished PhD Thesis; Reading University.

Khalifa, H. and Weir, C. (2009). *Examining reading. Studies in Language Testing* 29. Cambridge: Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95 (2), 163-182.

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

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

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

Kobayashi, K. (2009). Comprehension of relations among controversial texts: effects of external strategy use. *Instructional Science* 37, 311-324.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* Cambridge: Cambridge University Press.

Kroll, B. (1990). *Second language writing: Research insights for classroom.* Cambridge: Cambridge University Press.

Lacroix, N. (1999). Macrostructure construction and organisation in the processing of multiple text passages. *Instructional Science* 27, 221-233.

Langer, J. A. and Flihan, S. (2000). Writing and reading relationship: Constructive tasks. In R. Indrisano and J. R. Squire (eds.), *Perspectives on writing: Research, theory and practice* Newark, DE: International Reading Association.

Landauer, T. K., Laham, D. and Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In Shermis, M. D. & Burstein, J. (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. Hillsdale, NJ: Lawrence Erlbaum.

Masi, S. (2002). The literature on complexity. In L. Merlini-Barbaresi (Ed.), *Complexity in language and text* Pisa: Plus.

McNamara, D. S., de Vega, M. and O'Reilly, T. (2007). Comprehension skill, inference making, and the role of knowledge. In F. Schmalhofer and C. A. Perfetti (Eds.), *Higher level language processes in the brain* 233-251. Mahwah, NJ: Lawrence Erlbaum Associates.

McNamara, D. S., Louwerse, M. M., Cai, Z., and Graesser, A. (2005). Coh-Metrix version 1.4. Retrieved on 5 December 2008, from http//:Coh-Metrix.memphis.edu.

Messick, S. A. (1989). Validity. In R. L. Linn (Ed) *Educational measurement* 13-103. New York. American Council on Education. Mac Millian Publishing Company.

Meyer, B. J. F. (1999). Importance of text structure in everyday reading. In A. Ram and K. Moorman (Eds.), *Understanding language understanding* 227-252. Massachusetts: The MIT Press.

Moore, T. and Morton, J. (1999). Authenticity in the IELTS academic Module writing test: a comparative study of task 2 items and university assignments. *IELTS Research Report* 2, 64-106.

Moore, T., Morton, J. and Price, S. (2010). Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study. Unpublished Research Report, British Council/IELTS.

Munby, J. (1978). Communicative syllabus design. Cambridge: Cambridge University Press.

Nuttall, C. (1996). *Teaching reading skills in a foreign language.* Oxford: Heinemann.

Pallant, J. (2007). *SPSS Survival manual.* New York: McGraw-Hill.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Perfetti, C. A. (1991). Representations and awareness in the acquisition of reading competence. In L. Rieben and C. A. Perfetti (Eds.) *Learning to read: Basic research and its implications* 33-44. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Perfetti, C. A. (1997). Sentences, individual differences and multiple texts: Three issues in text comprehension. *Discourse Processes* 23, 337-355.

Perfetti, C. A., Britt, M. A. and Georgi, M. C. (1995). *Text-based learning and reasoning : Studies in history.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Perfetti, C. A., Rouet, J. F. and Britt, M. A. (1999). Toward a theory of document representation. In H van Oostendorp and S. R. Goldman (Eds.), *The construction of mental representations during reading* 99-122. London: Lawrence Erlbaum Associates Publishers.

Pressley, M., and Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale NJ: Erlbaum

Rayner, K. and Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, New Jersey: Prentice-Hall.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Rosenfeld, M., Leung, S. and Oltman, P. K. (2001). The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels. *TOEFL Monograph Series* MS - 21.

Rumelhart, D. E., and McClelland, J. L. (1986). (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press.

Rouet, J. F. (2006). *The skills of document use: from text comprehension to web-based learning*. New Jersey: Lawrence Erlbaum Associates, Inc.

Rouet, J. F., Britt, M.A., Mason, R. A. and Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology* 88 (3), 478-493.

Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: a multiple data source approach. *Language Testing* 17, 85–114.

Shiotsu, T. and Weir, C. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing* 24, 1-30.

Skehan, P. (1998). *A cognitive approach to language learning.* Hong Kong: Oxford University Press.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica: RAND Education.

Spivey, N. N. (1990). Construing constructivism: Reading research in the United States. *Written Communication* 7 (2), 256-287.

Spivey, N. N. (1991). Transforming texts: Constructive processes in reading and writing. *Written Communication* 7 (2), 256-287.

Spivey, N. N. and King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly* 24 (1), 7-26.

Spivey, N. N. and King, J. R. (1994). Readers as writers composing from sources. In R. B. Ruddell, M. R. Ruddell and H. Singer (Eds.), *Theoretical models and processess of reading* 668-694. Newark: International Reading Association.

Stahl, S A, Hynd, C R, Glynn, S M and Carr, M (1996a). Beyond reading to learn: Developing content and disciplinary knowledge through texts. In L Baker, P Afflerbach and D Reinking (Eds.), *Developing engaged readers in school and home communities* 139-163. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Stahl, S. A., Hynd, C.R., Britton, B. K., McNish, M. M. and Bosquet, D. (1996b). What happens when students read multiple source documents in history? *Reading Research Quarterly* 31 (4), 430-456.

Stanovich, K. E. (1991). Changing models of reading and reading acquisition. In L Rieben and C Perfetti (Eds.), *Learning to read: Basic research and its implications* 19-31. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E. (1996). Word recognition: Changing perspectives. In R. Barr, M.L. Kamil, P. Mosenthal, and P. D. Pearson (Eds.), *Handbook of reading research, Vol II* 418-452. New York: Longman.

Stromso, H. I. and Braten, I. (2002). Norwegian law students' use of multiple sources while reading expository texts. *Reading Research Quarterly* 37 (2), 208-227.

Stromso, H. I., Braten, I. and Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction* 20, 192-204.

Stromso, H. I., Braten, I. and Samuelstuen, M. S. (2003). Students' strategic use of multiple sources during expository text reading: A longitudinal think-aloud study. *Cognition and Instruction* 21 (2), 113-147.

Stromso, H. I., Braten, I. and Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction* 18, 513-527.

Taylor, L. 2007, The impact of the joint-funded research studies on the IELTS writing test, in *IELTS Collected papers: Research in speaking and writing assessment*, (Eds) Taylor, L. and Falvey, P. 479-492. Cambridge University Press: Cambridge.

TOEFL Official Guide (2005). Educational Testing Service. Princeton: ETS.

Urquhart, A. H. and Weir, C. J. (1998). *Reading in a second language: Process, product and practice.* Harlow: Longman.

Ünaldi, A. (2004). *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. (1977). Semantic macro-structures and knowledge frames in discourse comprehension. In M. A. Just and P. A. Carpenter (Eds.), *Cognitive processes in comprehension* 3-32. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

van Dijk, T. A. and Kintsch, W. (1983). The notion of macrostructure. In T. A. Van Dijk and W. Kintsch (Eds.), *Strategies of discourse comprehension* 189-223. New York: Academic Press.

van Oostendorp, H. (1996). Updating situation models derived from newspaper articles. *Medienpsychologie: Zeitschriff fur individual und Masseskommuikation* 8, 21-33.

Wainer, H., Bradlow, E.aT. and Wang, X. (2007). *Testlet theory and its applications.* Cambridge: Cambridge University Press.

Weigle, S. C. (2002). *Assessing writing*, Cambridge: Cambridge University Press.

Weir, C. J. (1983) *Identifying the language problems of the overseas students in tertiary education in the UK.* Unpublished PhD dissertation, University of London.

Weir, C. J. (2005). *Language testing and validation.* London: Palgrave MacMillian.

Weir, C. J., Hawkey, R. Green, T., Devi, S., Unaldi, A. (2009) *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, IELTS Research Report No.*9: 97-156. British Council/IDP Australia.

Wiley, J., Goldman, S., Graesser, A. C., Sanchez, C. A., Ash, K. A. and Hemmerich, J. A. (2009). Source evaluation, comprehension and learning in internet science inquiry tasks. *American Educational Research Journal* 46, 1060-1106.

Wiley, J. and Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology* 91, 301-311.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology* 83, 73-87.

Wolfe, M. B. and Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction* 23, 467-502.

Xue, G. and Nation, I. S. P. (1984). A university word list. *Language Learning and Communication* 3, 215–229.