

**Reading Comprehension Processes and Strategies in L1
and L2 in Malaysian Primary and Secondary Schools**

Volume I

by

Ismail Sh. Ahmad, B.A., M.Sc. Ed., Dip. TESL.



Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy
May 1997

Ph.D Thesis: An Outline of Chapters

**Reading Comprehension Processes and Strategies in L1
and L2 in Malaysian Primary and Secondary Schools**

Candidate: Ismail Sh. Ahmad
Supervisor: Professor Dr. Colin Harrison
Faculty: Education
University: The University of Nottingham

CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	(i)
SYNOPSIS <u>AN OVERVIEW OF THE WHOLE STUDY: THE PROBLEMS AND OVERALL AIMS</u>	1
 <u>PART I STUDY: A SMALL SURVEY</u>	
CHAPTER 1 <u>INTRODUCTION AND BACKGROUND TO THE PART I STUDY</u>	
1.1 Background/Overview to the Part I Study	5
1.2 Rationale of the Study	13
1.3 Guiding Parameters of the Study	16
1.4 Statement of the Problems	23
1.5 Initial Hypotheses	26
1.6 Importance of the Study	29
1.7 Definition of Terms	31
1.7.1 Comprehension	31
1.7.2 Holistic	33
1.7.3 Subskills	34
1.7.4 Bahasa as the First Language (L1)	35
1.7.5 English as a Second Language (L2)	37
1.8 Organisation of the Thesis	39
CHAPTER 2 <u>REVIEW OF LITERATURE PERTINENT TO THE PART I STUDY</u>	
2.1 Models of Reading Comprehension Processes in the First Language	
2.1.1 The Bottom-Up Model	42
2.1.2 The Top-Down Models	44
2.1.3 The Interactive Model	48
2.1.4 Rayner and Pollatsek's Bottom-up Interactive Model	49
2.1.5 Stanovich's Interactive-Compensatory Model	50
2.2 Models of Reading Comprehension Processes in the Second Language	52

2.3 Two Contrasting Theories in Reading Comprehension and Their Implications	54
2.3.1 The Subskills Reading Theory	55
2.3.2 The Holistic Reading Theory	60

CHAPTER 3 DESIGN AND METHODOLOGY OF THE PART I STUDY

3.1 Overview of the Part I Study	64
3.1.1 Choosing the Research Method (Descriptive)	66
3.1.2 Chronology of the Study	66
3.2 Research Procedure 1: Piloting the Comprehension Tests in Nottingham	69
3.2.1 Replicating Lunzer and Gardner's Eight Subskills: Piloting the Modified and Translated Four Texts in Nottingham	70
3.3 Research Procedure 2: Testing the Comprehension Subskills in Malaysia	74
3.3.1 Selecting and Testing the Four L1 and Four L2 Texts on the Primary and the Lower Secondary School Pupils	74
3.4 The Data Analysing Techniques	86
3.4.1 Factor Analysis (FA) and its Background. From Thurstone to Lunzer and Gardner's works	86
3.4.2 Performing FA and its Related Problems and Issues	90
3.4.3 Finding the Correlation Matrix by Item Analysis	94
3.4.4 Item Analysis: Interpreting its Correlation Coefficients	98
3.4.5 Principal Components Analysis (PCA)	100
3.4.6 Kaiser Normalisation (Oblimin Rotated Factor Matrix)	104
3.5 Finding the Reliability of the Eight Tests	105

	<u>Page</u>
CHAPTER 4 <u>RESULTS AND DISCUSSION OF THE PART I STUDY</u>	111
4.1 Presentation and Discussion on the Four L1 and the L2 Tests	
4.1.1 Item Analysis	111
4.1.2 Factor Analysis, Intercorrelation of Factors and the Eigenvalues	126
4.1.3 Distribution of Significant Loadings	137
4.1.4 Identifying the Pattern of Relationship	139
4.2 Discussion of Findings for the L1 and the L2 Tests	141
CHAPTER 5 <u>CONCLUSIONS FROM THE PART I STUDY AND ITS IMPLICATIONS FOR THE PART II STUDY</u>	145
 <u>THE PART II STUDY: MULTIPLE CASE STUDIES</u> 	
CHAPTER 6 <u>INTRODUCTION AND BACKGROUND TO THE PART II STUDY</u>	
6.1 Background and Rationale to the Part II Study	149
6.2 Parameters of the Study	152
6.3 Statement of the Problems	153
6.4 Hypotheses	154
6.5 Importance of the Study	155
6.6 Definition of Terms	
6.6.1 Average Students	156
6.6.2 Above Average Students	156
6.6.3 Comprehension Answering Strategies	156

	<u>Page</u>
CHAPTER 7 <u>REVIEW OF LITERATURE PERTINENT TO THE</u> <u>PART II STUDY</u>	158
CHAPTER 8 <u>DESIGN AND METHODOLOGY OF THE PART II STUDY</u>	
8.1 Overview of the Study	170
8.1.1 Choosing the Research Method	170
8.1.2 Designing the Multiple Case Studies	171
8.1.3 Chronology of the Part II Study	172
8.2 Research Procedure 1: Pilot Case Studies	174
8.2.1 Piloting in Nottingham	175
8.2.2 A Report From a Selected Case	177
8.3 Research Procedure 2: Testing and Interviewing the 64 Cases in Malaysia	
8.3.1 Selecting the Schools and the Two Groups of Students	181
8.3.2 The Design of the Interviews	184
8.3.3 Collecting the Data	185
8.4 Chronology of the Data Analysing Techniques	185
8.4.1 Transcribing the 64 Interviews	186
8.4.2 Translating the Chosen 16 Interviews into English and Finding the Reliability of the Translation from Bahasa to English	186
8.4.3 Coding the Discourse Units and Finding the Reliability of the Coding	187
8.4.4 Analysing the Coded Discourse Units for All the 16 Interviews	191
8.4.5 Presentation and Analysis of the Processed Discourse Data	192

	<u>Page</u>
CHAPTER 9 <u>RESULTS AND DISCUSSION OF THE PART II STUDY</u>	193
CHAPTER 10 <u>CONCLUSIONS AND RECOMMENDATIONS FROM THE PART I AND PART II STUDIES: SOME IMPLICATIONS FOR THE READING COMPREHENSION SYLLABUS, CLASSROOM PRACTICE AND FUTURE RESEARCH.</u>	232
10.1 Evidence of Unitary Subskills in Reading Comprehension from the L1 and L2 Tests and its Implications for the Reading Comprehension Syllabus and Classroom Practice	236
10.2 Findings and Implications of the Second Part of the Study for the Reading Comprehension Syllabus and Classroom Practice	248
10.3 The Importance of the Acceptability of Students' Comprehension Reasoning Processes and its Implications for Future Research	253
BIBLIOGRAPHY	261

Acknowledgements

My sincere appreciation to my supervisor Professor Dr. Colin Harrison for his total devotion, unstinting guidance and invaluable insights in advising and assisting me to accomplish this thesis. I am greatly indebted to him. I would also like to express my appreciation to Jane Restorick for the programming of the collected data. Her patience working with me is highly appreciated. I owe special thanks to Jill Seal for her patience, calmness and vigilance in assisting me in writing the thesis. I wish to record my grateful thanks to Pam Brown on the administrative aspects of my research progress.

My appreciation to the Malaysian Ministry of Education especially to the Educational Planning and Research Division, the local educational authorities in Johor Bahru, the affected headmasters and the teachers of the chosen schools. At the heart of this research are the affected students. Their willingness to respond to the comprehension tests and to the rigorous verbal interviews of the four reading comprehension texts are highly appreciated. Thank you to all the academic and non-academic staff of the Department of Education, which will soon become a Faculty of Education, and to the Research Centre of the International Islamic University of Malaysia (IIUM), the Public Service Department of Malaysia (PSD) and other related bodies directly or indirectly financing and supporting this educational study.

Above all, I am indebted to the trust shown by the IIUM, PSD and to the nation as a whole in granting me the scholarship to pursue and achieve this goal. My appreciation to my wife and daughters for their patience, continual support and understanding, without which the thesis would not have been completed.

Ismail Sh. Ahmad
May 1997

Reading Comprehension Processes and Strategies in L1 and L2 in Malaysian Primary and Secondary Schools.

Synopsis

This study is set in the context of the acknowledged debate, highlighted by the work of Lunzer and Gardner's Schools Council project (1979), concerning the theoretical issue of whether reading comprehension is a unitary competence or consists of identifiable discrete subskills. This long-standing polarised theoretical debate can be traced as far back as the sixties and seventies in the position taken by reading experts such as Spache and Spache (1969), Davis (1971) and Thorndike (1973). Spache and Spache and Thorndike concluded that reading comprehension was a unitary competence, not consisting of separate skills that can be practised in isolation. On the other hand, Davis viewed reading comprehension as composed of separate identifiable skills and abilities. The polarised arguments pose a question as to the nature of reading comprehension. Is there such a thing as discrete reading comprehension sub-skills that can be built up hierarchically and can promote the understanding of texts?

With the question in mind this study set out to test whether reading comprehension is a unitary competence of sub-skills or one that can be broken down into separate sub-skills. The research involved the rigorous testing of a series of reading comprehension tests in two languages using four texts taken from the work of Lunzer and Gardner (1979). The texts were modified to suit the socio-cultural context of the students. All of the chosen texts were translated into Bahasa (L1) which is the mother-tongue of the students.

In principle, the focus of the study in Part I is centered on replicating the work of Lunzer and Gardner (1979) in some selected Malaysian primary and secondary schools. It seeks to understand whether the main hypothesis holds that reading comprehension is unitary in nature and cannot be broken down into a number of distinct subskills. A selected 300 primary school pupils aged 12 were required to read and answer four comprehension tests written in L1. Another

selected 150 secondary school students aged 15 were required to perform the same tasks on material written in L2. Each test has about 30 comprehension questions which are divided into eight categories of subskills. The two groups produced a total of 1,636 valid comprehension tests which were marked rigorously. Factor analysing the data yielded a number of important findings concerning whether reading comprehension subskills are unitary or hierarchic in nature. These findings may suggest some recommendations for improving reading for learning across the Malaysian primary and secondary schools curriculum.

In Part I the outlined five chapters discuss the background information which led to the testing of the 450 students, the related literature review, the chosen research design and analysis, the findings and the research implications for the Part II study.

The study reported in Part II is an extension of the work done in Part I, in that the remaining five chapters explore the justification of conducting the in-depth interviews, the review of the related literature, the design of the interview, the findings and the educational implications of the study. This part explores the reading comprehension strategies that were used by the students in answering the comprehension questions. The second study was successfully made during the summer of 1994. A total of 16 students aged 15 were selected from several secondary schools in Johor Bahru, the capital state of Johor, Malaysia. Categorically, they represented the good and the average students in terms of overall academic achievement and reading comprehension. They were required to answer the four comprehension tests which were equally divided into L1 and L2 and subsequently each of them had to verbalise the reasons for the answers they had chosen or written. From the total of 64 interviews only 16 were chosen to be analysed. Each interview was rigorously transcribed and translated into English. Subsequent quantitative analyses were made to categorise the verbal protocol into the 8 categories of subskills. The findings of the analyses confirm the findings in the Part I Study and extend current understanding of the reading comprehension strategies used by the students.

The main objective of the Part II study is to explore the relationship between the eight types of question and the eight types of coded discourse which are based on the eight subskills of Lunzer and Gardner (1979). In general, each reasoning strategy used by the readers in response to each test item is interpreted and categorized as one of the eight subskills. Currently, no similar reading research has been done in this area.

In this respect, three general explorative hypotheses are put forward. First, there is a relationship in the distribution of the discourse units between the question types and the discourse types. This relationship must be established if one wants to proceed to the other two hypotheses.

Once the first hypothesis is established, the following hypotheses may proceed to seek a clearer understanding of the interaction between the so called "lower order" items and "higher-order" items through interpreting the verbal discourses of the answered questions across the two groups of readers. In other words, it specifically asks what sort of discourse types are used for each question type and what are the supporting or compensating discourse types that come hand in hand with each question type. Generally, it is speculated that both the good and the average readers may use the same subskills in responding to each test item, but percentage wise some differences may exist in terms of the loading of the discourse types due to the anticipated differences in the comprehension answering strategies of the two groups of readers. In other words, when responding to a particular test item a good reader may have a higher percentage of verbal input than an average reader. The reverse may also be true. An average reader may have a higher percentage of input for a particular test item than a good reader. Thus, performance wise, both groups may exhibit a variation of percentage differences across the eight subskills for each test item. The percentage differences are open to critical interpretations.

With this in mind, the second hypothesis is advanced: there is a difference between the good and the average readers in terms of the patterns of the discourse units. The results could cast some light onto the mental processes involved in both groups and extend our understanding of the nature of these

processes. It is also hoped that the results may give some ideas to reading teachers in assisting poor comprehenders to comprehend better. At this juncture it is important to note the nature of prior learning experiences in shaping different ways of solving a test item (Bloom et al., 1956. p. 16). In other words, each student in the two groups may choose different skills or strategies for each test item. But regardless of the prior learning experiences of the tested and interviewed readers, what matters in this study are the comprehension answering strategies of the readers and if prior learning experiences do have an important impact on the answering strategies then a higher percentage of the discourse type (*Forming Judgement*) that represents 'prior learning experience' is expected to emerge from the data analysis of the two groups of readers. Although the problem of 'prior learning experience' is not the main problematic area to be investigated, the results from the analysed data may reflect the importance of the problem. The problem is that it is not known whether the good readers reflect more on discourse type (*J*) than the average readers or whether the average readers use more of this skill than the good readers. As far as this problem is concerned, no similar research has been done in this area, especially in the context of the eight 'subskills'.

Finally, the third issue of Part II of this study advances the second hypothesis to another frontier: whether the different languages produce different patterns of responses across the eight subskills. In this regard, the final hypothesis is stated as: there is a difference between the average and the good readers in term of the distribution of the discourse units related to the language of the comprehension test passages and the language of the tested questions.

It was hoped that the results would shed light on the nature of reading comprehension in L1 and L2 and thus allow reading teachers and other interested groups such as curriculum planners and reading theorists to be better informed and equipped with relevant diagnostic tools, which could be developed from this study, to identify problems in reading comprehension in primary and secondary schools in Malaysia. The thesis ends with a discussion of the implications from both studies especially for the reading curriculum, instruction, pedagogy, classroom practice and future research.

Chapter 1

Introduction and Background to the Part I Study

1.1 Background to the Part I Study

Research in trying to understand the nature of reading comprehension is not conclusive and there has been a long-standing debate for more than half a century among psychologists, reading experts and educators. The findings of these scholars are generally divided into two contrasting groups; holistic and atomistic. Between these two groups is what is known as the intermediate group, which Rost (1989, p. 89) identified as follows: 'An intermediate position is presented by those studies which postulates only two components of reading comprehension, namely, 'vocabulary' or 'literal reading', and 'general language comprehension' or 'inferential reading'...'.

This study is concerned with the first two contrasting groups. The holistic group represents the view that reading comprehension skills are a unitary process (Smith, 1971; Goodman, 1967). In the *Review of Educational Research* , Dole and his colleagues (1991) summarised the historical origin of the subskills group in the contemporary American schools curriculum which dated back to the 1950s. Dole et al., (1991, p. 240) say:

The proliferation of comprehension skills and the comprehension curriculum as we know today emerged from this task-analytic behavioral conception of reading. Guthrie (1973) described this curriculum as an assembly-line model of skill acquisition. In such a curriculum, it is assumed that each skill can be mastered and that the aggregate of all the subskills equaled reading comprehension.

There were numerous attempts to identify discrete subskills in reading comprehension and the results are not conclusive. Lunzer and Gardner (1979)

acknowledged that the study by Davis (1968) was the most ambitious one. Davis (1968) carried out two 96 item tests on 988 college students and found that there are four separable skills. The four skills are known as ' (1) identifying word meanings, (2) drawing inferences, (3) identifying the writer's technique and recognizing the mood of a passage, (4) finding answers to questions.' (Lunzer and Gardner, 1979). The data and the findings were later scrutinised by other researchers including Spearitt (1972) who acknowledged Davis's findings but with a little modification to the fourth factor which he thought of as 'the ability to follow the structure of a passage'. In the United Kingdom Lunzer and Gardner (1979) carried out a landmark study on the nature of reading comprehension and found that reading comprehension cannot be broken down into distinct subskills and is far from being in a hierarchical order. Prior to the findings Lunzer and Gardner (1979, p.40) acknowledged that before the study was carried out, their '...weight of opinion inclined to the first hypothesis, that comprehension involves a multiplicity of subskills,...'. After factor analysing the data Lunzer and Gardner (1979, p.64) say:

We conclude that individual differences in reading comprehension should not be thought of in terms of a multiplicity of specialized attitudes. To all intents and purposes such differences reflect only one general aptitude: this being the pupil's ability and willingness to reflect on whatever it is he is reading.

This study is concerned with the debatable issues of distinct, hierarchical and unitary subskills in reading comprehension. The researcher observed that no similar or comparable study has been done in Malaysian primary and secondary schools particularly in testing reading comprehension with extensive texts and to be more precise in concurrently testing whether reading comprehension skills in the two languages are hierarchical, unitary or atomistic in nature. Thus, this study sets out to replicate and extend the Lunzer and Gardner (1979) study in a few selected primary and secondary schools in Malaysia. It is very important to note at the outset of this study that a replication study is bound to have some difficulties. Differences in the samples' cultural background, the objectives of the

educational policy of the affected countries and the syllabus content could have some impact as to the comparability of the results. This issue of the difficulty of doing comparative analysis across the many existing studies (in reading) is explained by Rost (1989, p. 89): 'Not only are they based on samples of different subjects and variables, but also the dimensions vary from one publication to another.' With this in mind, this study sets to minimise the anticipated differences by closely scrutinising and evaluating Lunzer and Gardner's study (1979). The greatest care has been taken in translating and adapting the original texts to the socio-cultural background of the readers.

Basically, the two languages taught and learned in the Malaysian schools are commonly known as Bahasa Malaysia and English. The former is also known as Bahasa Melayu which literally means 'Malay Language.' Out of the many languages in Malaysia, Haji Omar (1992, p. 2) says, 'Malay was no doubt the most predominant of these languages in terms of the number of speakers, geographical spread, and social as well as political status.'

There are two terms used throughout the study: L1 and L2. L1 is a term to connote the teaching and learning of Bahasa Malaysia in the schools. On the other hand L2 is a language taught and learned as a second language. As explained by Haji Omar L2 refers to the teaching of English as a compulsory subject in the primary, secondary and the tertiary levels (p. 11, 1992).

A background of the teaching of reading in English is noted by Ramaiah (1994, p. 79) who identified the changing trend that had taken place in the University of Malaya's shifting emphasis from teaching discrete skills of the 1970's bottom-up process to the teaching of integrated skills. This trend was noticable in the late eighties by the publication of the "Reading for Academic Study" to suit the needs of the Malay medium school leavers in coping with the reading materials which are mostly in English. This changing trend is also noted in the compendium used by the secondary English language teachers issued by the Kementerian Pendidikan Malaysia (1989a). The compendium reflects a communicative and integrated approach to the teaching of reading comprehension. Ideas like background knowledge, processing strategies at the

graphic, semantic, syntactic and discoursal levels, categories of reading skills such as skimming, scanning, guessing, predicting, inferring and varying reading speeds are emphasised (39-52).

In 1983, a new curriculum was introduced in all the primary schools in Malaysia. The curriculum is acknowledged to be vastly different from the old curriculum. Haji Sabran and Syam (1985, chapter 8) noted that the planning of the new curriculum in the teaching of L1 took into consideration many factors such as personal and societal development and they observed that the subject itself is taught with the view that language is a tool for communication. The content of the Bahasa Malaysia (L1) under the new curriculum is divided into two phases. The first phase is for years one to three and the second phase is for years four to six. (Haji Sabran & Syam, 1985, chapter 8). In general, the programme for the first phase emphasises the mastery of basic skills, and in the second phase, particularly in reading skills, the students are taught to understand, summarise and interpret information from a variety of sources.

This study had involved pupils who were at the end of their second phase of primary schooling. In the 'Special Guide Book for Bahasa Malaysia for Year Six' (Kementerian Pendidikan Malaysia, 1990), it is emphasised that the teaching programme of L1 should give more consideration to pupils' experiences and this is manifested through many activities such as telling stories, personal experience, famous people, contemporary affairs and culture. The Guide Book stresses:

In the content's selection, emphasis should be given to teaching and learning the language. The content must be chosen based on the lesson's objective and the skill that are to be delivered (all these) must be relevant to the pupils' ability and interest. The content of the teaching should have elements of citizenship.

(Translated, p. 17)

The Guide Book provides vital information in terms of the content, item, method, teaching and learning strategies and the main characteristic of the language

programme of the new curriculum is the idea of integration of language skills in one teaching and learning period.' (Haji Sabran & Syam, 1985, p. 91-92).

The integrated learning skills are designed in the best interest of the teachers and the learners. There are suggestions to the teachers as to the techniques of teaching and learning of the reading syllabus and these are in a hierarchical order; the lower-order skills, finding main points, understanding sequences of ideas or points, making predictions, drawing conclusions, generating questions and discussion. In short, the suggested techniques are a clear indication of the hierarchical view of reading comprehension and could be divided into literal, responsive and evaluative types of activities (Kementerian Pendidikan Malaysia, 1990). As an example, the comprehension exercise under the lower-order skills is suggested to the teachers in a hierarchical order. It begins with finding the meaning of new words from the dictionary or by discussion, finding the gist of the story according to the flow of the story, finding answers to literal comprehension questions and discussing the answers given by the pupils and various other activities that could foster understanding. The above techniques are only suggestions and teachers are advised to use their own creativity in choosing and applying the techniques that are thought to be relevant to the classroom environment and the pupils' achievements (Kementerian Pendidikan Malaysia, 1990).

The argument of the assumed hierarchical order skills in reading comprehension is not new. This hierarchical concept, also known as an implicational scale in reading comprehension, is what Anderson (1990, p. 425) comments on here:

...it is the experience of this author that it is common practice among teachers, testers and researchers of reading to assume that reading skills can be identified, taught, tested and researched. It is furthermore common for reading specialists to refer to lower and higher order skills, implying both a hierarchy of such skills, and an implicational scale (or cumulative hierarchy), such

that lower skills are held to be necessary before higher ones can be acquired or developed.

This implicational scale or multiplicity of subskills is therefore thought to exist in the planning of the reading comprehension syllabus for the Malaysian primary schools. It is also assumed that at the end of the year the pupils in year six should have learned the reading comprehension skills as proposed in the syllabus.

In the English language (L2) syllabus content for the primary school the concept of reading skills is stated as follows:

A hierarchy of reading skills is drawn up and linked laterally with the aural-oral skills and writing. The aim is to provide the pupils with the basic reading skills of word recognition and phonics and then to lead them on to comprehension skills. This reading component will also provide the pupils with the opportunity of developing study skills such as using dictionaries and encyclopedias and reading maps, plans and graphs...All the skills listed under the various headings, are not necessarily in hierarchy. The skills and functions are stated in terms of pupil performance.

(Kementerian Pendidikan Malaysia, 1981, p. 2)

What can be noted from the above quotation is that the concept of a hierarchy is assumed to be unproblematic. This assumption is of a fundamental part of the understanding of reading and the view of the reading curriculum expressed is one of the fundamental reasons in conducting the study. This study is questioning the hierarchical view inferred from the above quotation. Although it is indicated in the above syllabus that not all skills are necessarily hierarchical it is still set in the view that reading is hierarchic in nature. In the *Conpendium: A Concise Guide for Teachers of English* (Kementerian Pendidikan Malaysia, 1989a, pp. 38-52), reading skills for the secondary school are divided into six categories: skimming, scanning, guessing, predicting, inferring and varying

reading speed. These categories, including intensive and extensive reading, are some of the ways or techniques into reading (Grellet, 1981). In the compendium the reading skills are also described as comprising of four different levels of processing strategies: graphic, semantic, syntactic and discoursal. The graphic level is thought to be the least difficult and the discoursal level is considered to be the most difficult for students to learn. Although the guide offers many practical ideas and suggestions to language teachers, it represents the hierarchical view of reading skills and processing strategies.

Research on whether it is possible to identify specific skills, which are believed to be built up hierarchically in the reading comprehension of secondary school children in English, was done by Lunzer and Gardner (1979) in the Schools Council Effective Use of Reading project. They concluded that reading comprehension is a 'unitary aptitude'. Another pilot study by Rost (1989), in trying to better understand the unquestioned truth of the assumed trainable 'subskills' in 220 German second grade elementary school children, found the following:

So far there is no sufficient reason to assume that
several clearly distinguishable and (from the point of
view of a psychologist or reading specialist)
meaningfully interpretable subskills of reading
comprehension exist among second grade pupils. (p. 106)

This unitary nature of reading comprehension processes is also shared by Wallace (1992) as:

...similar processing strategies in the reading of all
languages, even when the the writing systems are
very different. (p. 22)

In questioning the assumption that reading comprehension skills are hierarchical in concept and after reviewing the design of the reading skills syllabus for both the L1 and L2 which reflect this assumption, this study set out to replicate

Lunzer and Gardner (1979)'s research in selected primary and secondary schools in Malaysia.

The essence of this study is to explore in greater depth than was possible in the Lunzer and Gardner (1979) study, and with a particular emphasis on the L1 and L2, the issues of whether one can identify a cumulation of distinct subskills in reading comprehension or whether reading comprehension subskills are unitary in nature. In doing so, the researcher has replicated and extended the work carried out by Lunzer and Gardner (1979) in England. Lunzer and Gardner (1979) tested a total of 257 primary school children aged 10-11 years who were tested by four reading comprehension tests in the English language.

The same tests were translated into Bahasa Malaysia (L1) and later tested on 271 primary school pupils aged 12. In Malaysia compulsory education starts at the age of seven. The tests were also used in English (L2) on 150 secondary school students aged 15. Comparatively this study replicates the Lunzer and Gardner (1979) study in terms of the age group and the main language used in the classroom. It also extends the study by testing the original English texts on the secondary school students. It must be noted that the tested pupils in Malaysia started their compulsory education between the age of 6 and 7 and in general this is about a year behind the pupils in England. What is important is the fact that the tested primary pupils in both countries had received an equal number of years of formal education especially in the context of learning to read. All of the primary pupils had been in school for at least six years.

It is also important to note the fact that the original texts were carefully translated and adapted to the socio-cultural background of the targeted pupils without changing the gist, flow and content of the stories. It is also acknowledged that in the process of translating and adapting the texts to the prior knowledge of the pupils too many alterations to the original texts may invalidate a direct comparison with the original study. So, such modifications are kept to the very minimum.

1.2 Rationale of the study

The fundamental reason for conducting this study is to investigate whether reading comprehension in L1 and L2 is a unitary competence or a multiplicity of subskills in selected primary and secondary schools in Malaysia. Along with the above main objective is the hope that the findings of this study may be useful in:

- a) enhancing the understanding of reading comprehension in L1 and L2.
- b) producing a better framework for teaching and assessing reading comprehension. This may include providing better perspectives in understanding reading comprehension to pre-service teacher-trainees, curriculum designers, textbooks writers, parents and other related bodies.
- c) assisting and equipping in-service teachers with progressive concepts in reading comprehension and its assessment.
- d) assisting the Ministry of Education (MEO) or commonly known as Kementerian Pendidikan Malaysia (KPM) in raising the standard of reading in the primary and secondary schools and achieving the target of 100% literacy rate at the school level by the end of the century. The findings of this study may be useful in assisting the KPM and the Curriculum Development Centre (CDC) of KPM in planning the reading syllabuses in L1 and L2 and
- e) improving our understanding of children's reading comprehension. Oakhill (1993, p. 63) observed that the majority of research done in the area of children's reading has something to do with *words* and 'Far less work has been done on children's reading *comprehension* and how it might be improved.' This study sets out to build the researcher's knowledge foundation on how reading comprehension in the primary and secondary could be improved.

Apart from the above logical bases, this study is also triggered by the UNESCO finding on the literacy rate in Malaysia. Mustapha (1994) stressed that the current literacy rate in Malaysia is 78.5%. Although this rate is high, it is far from the MEO target of 100% literacy rate. In 1993, the Federal Inspectorate of Schools conducted an extensive study on Remedial Education. A total of 879 primary schools were selected. The gist of the findings is highlighted in the fact that '...9.8% of the pupils in the sample schools were reading and 7.3% were writing at levels considerably below those of their peers in the language of instruction.' (Mustapha, 1994, pp. 1-2). Again, these disturbing statistics on reading and the possible factors contributing to them are worth investigating.

Commenting further on the rationale of the study are the alarming findings of a survey made on the teaching of reading comprehension in L2 in Malaysian schools. According to Mustapha (1994) a national survey was launched in 1983 after a small survey was done by the Federal Inspectorate of Schools in the state of Selangor in 1982. In this small survey Mustapha (1994, p. 5) makes an important quotation on the definition of comprehension skills which is defined in the survey as 'the ability to decode the message by drawing on syntactic and lexical clues.' Although no further explanation is given on this broad definition of comprehension skills, it is thought that it could be a reflection of the bottom-up model of the reading process and its approaches. The findings from the survey highlighted the need to facilitate English teachers with proper knowledge and understanding of the aims and procedures in teaching reading comprehension.

It is also noted that in all the reading comprehension lessons observed during the survey were three common practices performed by the teachers; focusing the pupils on reading-aloud, improving vocabulary and pronunciation and focusing on the product rather than the comprehension process.

The follow-up national survey in 1983 was an extension of the small survey. It involved a questionnaire and open-ended questions on some 234 lessons performed by the lower secondary school teachers in 61 rural and urban schools. Among the findings of the national survey was the fact that about a quarter of the samples were judged to be less than satisfactory. It is also noted

that many of the reading texts did not consider important factors such as the readability level, students' interest and the suitability of the texts to the cultural values of the readers. The majority of the teachers failed to realise the importance of other teaching qualities such as preparatory work, classroom participation, motivation, purposeful questioning, teaching aids and the need '...to help students develop proper reading strategies and acquire comprehension skills necessary for processing information in reading texts.' (Mustapha, 1994, p. 7).

Another related rationale for conducting this study is a survey done almost a decade ago in the state of Terengganu, Malaysia. A total of 58 primary schools were examined and evaluated in terms of the teachers' approaches, methods and techniques in teaching reading comprehension skills and writing in L2. In this survey the members of the English Panel of the Inspectorate reported similar findings in that '...the visit confirmed that comprehension skills were not systematically developed in most of these schools in Terengganu...that in the teaching of writing the pupils were not provided with sufficient opportunities to record or to communicate their ideas...Writing assignments were so highly controlled and form-focused that little was done to develop and enhance the pupils' ability to write creatively and elegantly.' (Mustapha, 1994, pp. 7-8). These negative findings especially in the teaching and learning of reading comprehension skills in Malaysian schools merits further proper investigation.

In doing so, this study begins by studying and testing the fundamental classical issue of whether reading comprehension in L1 and L2 consist of acquiring a set of discrete subskills, which many believe are in a hierarchical order, or whether it cannot be broken down into discrete subskills.

1.3 Guiding Parameters of the Study

Samuels and Kamil (1984) highlighted the fact that in building any good model one should consider three important characteristics; a good model should be able to provide as much information as possible as to previous and current ideas as well as projecting hypothetical testable constructs. In building up the following three models in reading comprehension, past and present related models are taken into account.

Although it is difficult if not impossible to make a perfect comparison of the contemporary models of the reading process, mainly due to '...lack of common focus...among the models [and] Model evolution over time...' (Samuels and Kamil 1984, p. 220), it is still imperative to understand as much concepts as possible from the different reading models particularly the bottom-up, top-down and interactive models (see chapter 2 for the review of the related literature). In the light of this study it is important that some implied hypothetical reading models be constructed. The models could serve as the guiding parameters of the study.

The characteristics of the study and its parameters are set in the context of three implied models of reading comprehension subskills; unrelated subskills, hierarchy of subskills and unitary subskills. Each model implies three different performances that are thought to exist among good, average and poor students. The yardsticks used in judging probable performances among the students are the eight putative subskills that are assumed to comprise comprehension. These eight categories of subskills are replicated from the work of Lunzer and Gardner (1979). The eight categories of comprehension questions are built in an hierarchical manner and degree of difficulty in order to suit the age-level of the pupils and to find the existence of a hierarchy of subskills in reading comprehension. Lunzer and Gardner (1979, p. 45) claim:

The present categorization was conceived of as partly hierarchical (cf. L, ISS, IMS), and partly corresponding to very clear differentiations (cf. W, WIC, M). If there are different subskills,

then a categorization like the present one is more likely to uncover them.

In this study, the definition of the eight subskills are taken from Lunzer and Gardner (1979, p. 44) and Waite (1980, pp. 21-22). They are:-

1. *Word meaning* (W) - In isolation.
2. *Words in context* (WIC) - Deriving the appropriate meaning of an ambiguous word from the context in which it appears.
3. *Literal comprehension* (L) - Finding the answers to questions when these can be obtained directly by reference to a phrase or a sentence in the text.
4. *Drawing inferences from single strings* (ISS) - A *string* is an uninterrupted sequence of words, usually a phrase or a short sentence. Questions in this category require the reader to draw an inference from such a sequence as opposed to deriving its literal meaning.
5. *Drawing inferences from multiple strings* (IMS) - These tasks are similar to ISS, save that the necessary information for making the inference cannot be found by reference to one phrase but must be deduced from a comparison of two or more facts appearing in different parts of the text.
6. *Interpretation of metaphor* (M) - These questions require the reader to show an understanding or appreciation of meanings that are given indirectly by use of metaphor.
7. *Finding salients or main ideas* (S) - The ability to isolate the key points of a passage.

8 *Forming judgements (J)*

- This category was originally thought of under the heading of 'evaluation'. However, while the items clearly require the reader to go beyond the text, he is not asked to make a value-judgement about the worth-whileness of the passage or of its presentation but rather to offer an intelligent interpretation of ideas contained in the text or implied by it in the light of his own knowledge of related matters.

The eight categories of comprehension subskills serve as the guiding parameters of this study. With such parameters, three hypothetical models of reading comprehension subskills were carefully designed and served as the guiding tools in the design of the study and in the interpretation of the factor analysis. The three models and their explanations are given below. It is also important to note that the models are built on the probable hypothetical behaviour of the written responses; being right or wrong in the light of the predetermined answers. Bear in mind that the models are also usable in interpreting the verbal protocol in the Part II of this research.

Model 1: Unrelated Subskills

In this model the poor readers are assumed to be less good than the good readers; so there are a few correct responses (pluses) and the pluses could be observed anywhere along the eight assumed subskills. There is no clear-cut pattern of responses and the responses are not related. The average readers just have a few more pluses than the poor readers. The good readers have more pluses than the poor and the average readers. All the responses across the three groups are not in any particular order.

Model 2: Hierarchy of Subskills

In this model the poor readers may be good at '*Word Meaning*' or '*Words in Context*' or with a few pluses on '*Literal Comprehension*' but as the questions climb up the hierarchy of difficulty the rest of the responses are thought to be dominated by bad or wrong responses. The average readers are expected to be a little better than the poor readers with a few more pluses. The good readers are thought to be good in all of the subskills. It is also important to note that no one is going to be good at the highest level, the '*Forming judgements*' questions, without being good at all those 'assumed' subskills lower than the '*Forming judgements*' questions.

This is what known as the hierarchy model. In it, once the pluses stop at any level, then it's assumed that the readers could do any question below the pluses or up to the maximum pluses achieved and not any more above. In other words, each subskill is accumulated from the assumed "lower-order" to the "higher-order" skills and in this case answering the '*Forming judgements*' questions is assumed to be problematic without mastering the other lower subskills; W, WIC, L, ISS, IMS, M and S.

Model 3: Unitary Subskills

This model implies that poor readers are going to be bad at most things; too few pluses. It can be assumed that not knowing too many words, failing to understand the concepts and the contexts of the story may actually hinder comprehension and thus affect the responses. The average readers are going to be fairly good at every skill. The good readers can be good at every skill. What is predicted by this model is a combinative-interactive of various subskills exhibited by the three group of readers.

The above models outline several general assumptions of what thought to exist out of the students' written performance of the comprehension tasks. So, the general inferred concepts explained formed the basis for investigating the students' performances in the light of L1 and L2 texts. The concepts of the three models are

best illustrated in the following Table 1 :

Table 1 : Models of Reading Comprehension Subskills

	Model 1 Unrelated			Model 2 Hierarchy			Model 3 Unitary		
	P	A	G	P	A	G	P	A	G
Subskills									
J	-	+	+	-	-	+	-	~	+
S	+	-	+	-	-	+	-	~	+
M	+	-	-	-	-	+	-	~	+
IMS	-	+	+	-	-	+	-	~	+
ISS	-	-	+	-	+	+	-	~	+
L	+	+	+	~	+	+	-	~	+
WIC	-	-	-	+	+	+	-	~	+
W	-	+	+	+	+	+	-	~	+

Note: The letters 'P', 'A' and 'G' stands for Poor, Average and Good students respectively. The negative signs '-' mean wrong answers. The positive signs '+' mean correct answers and the ~ signs indicate that the '+' and the '-' answers are somehow thought to be equally observed.

Another critical aspect of this study is the fact that the study is a replication of the work of Lunzer and Gardner (1979) through the School Council's *The Effective Use of Reading* team on the issue of whether reading comprehension comprises a set of discrete subskills. Lunzer and his team tested four different comprehension tests on some 257 pupils aged 10-11. They found no evidence that reading comprehension constitutes a set of discrete subskills. What is important in this study is the idea of not only replicating Lunzer and Gardner's (1979) study on the hypothesised subskills to the same age group in Malaysia but also extending it to a selected group of students aged 15. What is unique about this latter group is that they are given the original comprehension texts and tests as used by Lunzer and Gardner (1979) without translation into the Bahasa or Malay language.

The texts are slightly altered to accommodate the sociocultural background of the students. This is done in view of a relevant study on The Effect of Content Familiarity of the Comprehension of Text by Learners of English as a Second Language by Safiah (1985). Safiah selected 160 form four upper secondary native Malay students from two schools in the state of Selangor, Malaysia. The students had to sit four comprehension tests which were equally divided into two topics: wars and festivals. A familiar and unfamiliar passage were prepared for each topic. A comprehension test was conducted for each passage and in each test there were eighteen multiple-choice comprehension questions. The eighteen questions were equally divided into the literal, inferential and evaluation types of questions. Safiah found that students comprehend better when given familiar texts that suit the students' prior knowledge than when tested on unfamiliar ones. Safiah also stressed the importance of prior knowledge in the understanding of the familiar passages than the unfamiliar ones. Interference in comprehension due to the mismatch of readers' cultural background knowledge with L1 and L2 texts are also documented in Steffensen (1987).

In this study, the author took considerable measures in terms of the readability and suitability of the texts. A consensus was agreed upon after a series of consultations with the headteachers and the language teachers of the chosen schools. Their knowledge, views and experience in L1 and L2 suggested that the texts were suitable for the selected age group. Prior to the consultation, three

selected students of different language proficiencies pilot-tested the texts and it was found that in general the texts were appropriate to their second language proficiency.

1.4 Statement of the Problems

Research in understanding the complexity of reading particularly in reading comprehension is not new. A review on comprehension research by Venezky (1984) reports the first study on reading comprehension by Romanes in 1884. Romanes observed the behavioural differences in reading rates and written recall from a short ten-line paragraph tested on adult readers. In 1968, Davis carried out the seminal detailed study on the issue of whether there are any discrete subskills in reading comprehension. Davis (1968) claimed that he could identify four skills; identifying word meanings, drawing inferences, identifying a writer's technique in creating the mood of the passage and finding answers to questions.

Contrary to Davis's findings is the conclusion drawn from the work of Lunzer and Gardner (1979) that reading comprehension does not comprise a set of discrete subskills. Assessing the research findings of Davis (1968) and Lunzer and Gardner (1979) and all the complexities involved in their research has not been easy. In general both studies used different samples and texts and have reached different conclusions. Subsequent attempts have been made to reanalyse Davis's data (refer to chapter 2).

The research of Lunzer and Gardner (1979) provides a driving force for replicating the study in Malaysia. In a sense, this research queries the nature of reading comprehension in both L1 and L2 in some selected primary and secondary schools in Malaysia. It also seeks to indirectly understand the relationship between the main issue of whether reading comprehension is comprised of a hierarchy of discrete subskills or whether the subskills are unitary in nature, and the issue of the effectiveness of teaching reading comprehension in primary and secondary schools in Malaysia. Mustapha (1994) highlighted the findings from a national survey on the teaching of reading comprehension in Malaysia (see Rationale of the Study). The survey highlighted the lack of effectiveness in teaching reading comprehension and other related matters. Tying the findings as observed by Mustapha (1994) with this research is of paramount importance. This study and the findings from Mustapha (1994) may be useful in

projecting some possible related teaching and learning issues such as:

1. The instructional theories and principles as practiced by teachers in primary and secondary schools in teaching reading comprehension: a hierarchy of skills, unitary skills or a combination of both.
2. What are the common practices or beliefs or preconceptions notions of those teachers in the classroom in helping to increase the students' ability to comprehend the text?
In other words, what are the methods or techniques (before, during, and after reading) practised by the teachers and could such methods improve students' comprehension of the text? Are the teaching approaches (theoretical foundations), methods and techniques guided or influenced by the subskills and the unitary paradigms or are they eclectic ?
3. Are the teaching approaches in reading comprehension based on research findings or simply based on unquestioned truth/belief?
4. Is there a multi-dimensional standard framework practised for an effective diagnosis and assessment of reading comprehension in L1 and L2 in primary and secondary schools in Malaysia?

Any attempt to understand the complexities of teaching and learning reading comprehension is bound to confront the extent of the teachers' and students' perceptions as to what constitutes reading. In relation to the perceptions are the approaches carried out by the teachers in evaluating the students' performance in reading comprehension. It is important to stress that this study is trying to understand such complexities by understanding the popular conventional approach as practised by the teachers. This conventional approach regards reading from the behaviourist paradigm '...as a hierarchy of specific skills, a taxonomy of behaviours, which, if taught in small enough units, one upon the other, with rewards for each demonstration of success, could be built up into a

total edifice of reading ability' (Waterland, 1988, p. 11). In a like manner, Alderson (1990, pp. 425-426) notes that it is '...common for reading specialists to refer to lower and higher skills, implying both a hierarchy of such skills, and an implicational scale (or cumulative hierarchy), such that lower order skills are held to be necessary before higher ones can be acquired or developed...the impression of the author [is] that most practitioners assume both a hierarchy and an implicational scale.'

From this popular conventional approach a main research problem is raised: Is it possible to identify a distinguishable hierarchy of structured subskills in reading comprehension in L1 and L2 in the selected primary and secondary schools in Malaysia? On the one hand, if it is identifiable, how reliable or robust are the patterns of the hierarchy? On the other hand, if it is not identifiable, other meaningful theories in teaching and learning reading comprehension and its assessment paradigms should be suggested and forwarded to the teachers, related educational bodies, curriculum makers and textbooks publishers.

1.5 Initial Hypotheses

In relation to the statement of the problems discussed earlier, it is best to empirically study the subskills issue by assessing the products of the students' comprehension through a comprehension assessment. This could yield some ideas and strategies to researchers, curriculum-makers and teachers in improving reading comprehension. In the case of the subskills issue Johnston (1983., p. 5) says:

It is useful to know of a set of remediable subskills which comprise reading comprehension. This would provide a framework for effective assessment and remediation of reading comprehension difficulties. If we are to locate such a set, we must conduct a search that is driven by theory rather than by solely pragmatic concerns. Thus we must begin to elaborate a coherent theoretical model of reading comprehension.

With such a view a hypothetical research question is brought forward in the light of the empirical work of Lunzer and Gardner (1979). In their study the main hypotheses are stated as follows:

...reading comprehension can be broken down into a number of distinct subskills or, alternatively, that it constitutes a single aptitude or skill, one which cannot usefully be differentiated...(and)...some pupils might 'possess' lower-order skills but not higher-order skills. (pp. 63-64).

Lunzer and Gardner (1979) tested 257 children , aged 10-11 years, on four reading comprehension tests. The representative samples were chosen from four primary schools in Nottinghamshire and Derbyshire. A battery of four reading comprehension tests, each test consisted of about thirty questions which were arranged in an hierarchical order of difficulty. The questions were divided into

eight hypothesized subskills. They found no evident of a hierarchy of skills or individual differences in 'multiplicity of specialized aptitudes' and concluded that individual differences in reading has to do with 'ability and willingness to reflect' and recommended that a good comprehension test should include a variety of tasks that stimulate reflection (pp. 64-69). In the light of the preceding discussion, the main question of this study is best stated as:

Does Lunzer and Gardner's (1979) hypothesis that reading comprehension is a unitary competence hold for readers in an L1 and L2 context in selected primary and secondary schools in Malaysia?

This main question has also led to many hunches as to the answers of the question. Writing the hunches in an acceptable way has led the researcher to simplify the main question into a statement of possible outcomes of the study. Below is a comparison between the initial hypotheses used by Lunzer and Gardner (1979) and the ones projected in this study:

Lunzer and Gardner (1979, p. 63-64)

'...reading comprehension can be broken down into a number of distinct subskills or, alternatively, that it constitutes a single aptitude or skill, one which cannot usefully be differentiated... In the course of this inquiry, a third alternative (hypothesis) was examined, that some pupils might 'possess' lower-order skills but not higher-order skills...'

What Lunzer and Gardner (1979) found was that the factor analysis data strongly supported the second hypothesis.

Hypothetical constructs of this study:

Reading comprehension subskills are unitary in nature and far from being in any cumulative hierarchy in both L1 and L2 in some selected primary and secondary schools in Malaysia.

Such hypothetical constructs are very important in clarifying the overall objectives of this replication study. Although these are different hypothetical constructs from Lunzer and Gardner's (1979) study, the essence of both studies is still the same in that both studies seek the existence of subskills in reading comprehension. The present study explores the possibilities of the hypotheses in two languages in Malaysian schools: Bahasa (L1) and English (L2).

1.6 Importance of the Study

At the outset of this study is the hope that a wider understanding of the nature of the 'subskills' issue in reading comprehension could be shared among many interest groups such as reading researchers, curriculum-designers, teachers, reading specialists in the local educational authority, teacher-training colleges and universities and others who are directly and indirectly involved in improving the teaching of reading comprehension. It is hoped that the findings from this study will give some useful practical suggestions, guiding principles and opportunities to the above parties in designing teaching and learning methodologies for reading comprehension. It is also hoped that the findings from this study will suggest a better evaluation procedure for reading comprehension performance.

In the light of the above discussion, this study indirectly highlighted the need to review the teaching and learning of reading comprehension that emphasised discrete and hierarchically structural subskills. It is known that the independent sequential development of a set of hierarchically subskills' view is being challenged by a current view that emphasised active comprehension processes on the part of the readers. In other words, '...reading is now viewed as an active process in which readers select from a range of cues emanating from the text and the situational context to construct a model of meaning that represents, to some degree, the meaning intended by the writer.' (Dole et al., 1991, p. 255). Thus, in reading comprehension what readers say or write may be better understood if we give some attentions as to *how* they understand the text. In other words, active comprehension processes may involve active interdependence and interaction of the skills and encompass variables from within and outside the reading materials. It is important to learn that in traditional comprehension skills teaching, such as the eight most likely 'sub-skills' used by Lunzer and Gardner (1979), learners are expected to use the sub-skills in finding the intention of the author. Consequently, the evaluation could not escape testing the readers' acquisition of such skills. Thus, the readers are producing the products of acquiring the skills and the danger with the product of the evaluation is a tendency to view successful comprehension only from the perspective of the products of the reading performance.

Such traditional evaluation needs a better approach for understanding and evaluating the reading comprehension processes involved in arriving at products or answers, especially with regard to how the readers use the skills in monitoring their own comprehension strategies. In a way, in order to understand and to evaluate the reading comprehension processes, ones need to understand the cognitive view of reading comprehension such as the idea of the interactive nature of reading comprehension (Rumelhart and Ortony, 1977) that encompasses the prior knowledge that the readers bring into the reading tasks (Anderson and Pearson, 1984) and the use of strategies in scaffolding their own understanding (Dole et al., 1991).

Part I of this study highlights the issue of the discrete and hierarchical nature of subskills in reading comprehension. If the assumptions that reading comprehension subskills are not discrete or in any hierarchical pattern are proven, then popular reading programmes such as the traditional basal reading programmes need better ways in developing better ideas that can facilitate awareness in readers of how to comprehend various reading tasks across various reading genres. Such reading awareness may be enhanced by teaching comprehension strategies (Dole et. al., 1991) because such strategies, which can be effectively taught, may enhance the readers to actively interrogate the text in striving for meaning. In a way, the reading strategies as advanced by Dole should not be viewed as a hierarchy of 'sub-skills' but of a more active and conscious attempts on the part of the readers in making sense of the text read. In a way, in reading comprehension assessment, consideration should be given in how a reader uses the reading comprehension strategies because relying on the traditional methods of reading comprehension assessment gives little information about the processes through which a reader arrives at such understanding.

1.7 Definition of Terms

The following terms are often used throughout this study. It is important that proper definition and commentary are given to each specific term. It is hoped that the attempt is worthwhile in avoiding any misunderstanding that may arise from this reading research.

1.7.1 Comprehension

The term 'comprehension' has been defined and redefined by many reading scholars, from a simple common definition of comprehension as 'understanding what is read' as observed by Stierer and Bloome (1994) to the numerous attempts to identify the subskills involved in reading comprehension (Davis, 1968; Spearitt, 1972; Thorndike, 1973; Lunzer and Gardner, 1979). For instance, about two decades ago Chall (1977) observed two polarised views in the theoretical debate of reading comprehension: whether reading comprehension is a general skill or comprising specific identifiable skills. The former view is supported by the work of Davis (1971) who viewed comprehension as '...composed of separate skills and abilities, such as understanding word meanings, verbal reasoning, getting the main idea, detecting the author's mood, and discerning word meanings in context.' The latter is supported by Spache and Spache (1969) who viewed reading as a 'total act'.

Thorndike (1973) expressed the same position as Spache and Spache (1969) and viewed reading comprehension as a single aptitude but mainly consisting of verbal reasoning. It is important to learn the historical development of the teaching of reading that shaped the definition of the term. Stierer and Bloome (1994) stated that in the United Kingdom and in the United States the trend in the teaching of reading was an emphasis on pupils reading a series of passages of increasing difficulty. This idea was manifested in the popular McGuffey readers of the mid 1800s and the early 1900s and the basal readers and reading schemes from the 1940s to the 1970s. In the 1960s, reading research began to be viewed from many disciplines including cognitive psychology, linguistics and psycholinguistics.

What is notable in the focus of the various disciplines was the emphasis on comprehension of the text and subsequently this materialised in many reading tests, basal reading programmes and reading schemes (Stierer and Bloome, 1994). The shifting emphasis from oral reading to 'what is going on in the mind' was not without arguments among educators as to what to focus on, especially in early reading instruction. Comprehension is also understood as getting the meaning in the text and this is a familiar concept in many reading programmes where the pupils are taught to find the answers to a set of questions by first reading a passage or a short text. The right and wrong answers to the questions are already set by the teachers (Stierer and Bloome, 1994).

Another view emerged with the emphasis that meaning does not reside in the text itself. In contrast to the earlier view, reading is thought of as an interactive process between the reader and the text and thus comprehension could be understood as constituting a different interpretation for each different reader because of prior background knowledge of the reader and the context or purpose of the reading (See chapter 2). The current focus of redefining comprehension in research and education is much more than from the psychological perspective. It has embraced multiple perspectives from the field of literary, anthropology and sociological theorists among others. Stierer and Bloome (1994, p. 33) note:

Some literary theorists have suggested that there is no meaning *per se* and that all meanings can be deconstructed into nothingness. Some sociological theorists (often referred to as poststructuralists) have suggested that meaning is inherently unstable, incomplete and always political.

The comprehension process is also viewed in term of the complex relationship between vocabulary and comprehension (Ruddell, 1994). Ruddell (1994) suggests that the reader's prior knowledge and previous experience, information in the text read, reader stance in relationship to text and social interaction have been recognized as influencing vocabulary development. Ruddell's (1994, p.

415) definition of comprehension is stated as '...a process in which a reader constructs meaning while, or after, interacting with text through the combination of prior knowledge and previous experience, information in text, the stance he or she takes in relation to the text, and immediate, remembered, or anticipated social interactions and communication.'

In order to avoid confusion from the various definitions of comprehension, it is pertinent that throughout this study the term 'comprehension' follows Ruddell's (1994) definition and the definition taken from Lunzer and Gardner (1979, p. 38):

'...to penetrate beyond the verbal forms of text to the underlying ideas, to compare these with what one already knows and also with one another, to pick out what is essential and new, to revise one's previous conceptions.'

1.7.2 Holistic

Historically, two main views in the development of teaching reading proficiency have been the holistic and the subskill approach (Samuels, 1980). The former represents a whole-to-part idea as opposed to the part-to-whole view of the latter. The underlying ideology of the term in relation to the teaching of reading is that comprehension should be the main focus of learning and entails the use of multiple skills in the quest for meaning (Goodman, 1982). With regard to the purpose of teaching and learning sequences of skills, Goodman (1982), a psycholinguist, views reading as a natural language process and not as a set of skills to be learned. Goodman (1982) says:

No researcher has been able to support any particular sequence of skill instruction as having any intrinsic merit which derives from linguistic or psycholinguistic analysis. All sequences are arbitrary, often frankly stated

to be so by those who originally formulated them, though
not always by later borrowers. (p. 48)

1.7.3 Subskills

This term is interchangeably used by Lunzer and Gardner (1979) with another term 'skills' to mean the same thing; referring to the hypothesised hierarchy of eight subskills. The correlational studies of reading comprehension skills by Davis in the early 40's and the subsequent reanalysis of the data by Davis (1968) and Spearitt (1972) make use the term 'subskills' to mean 'skills'. Interestingly enough, Rosenshine (1980) found that some researchers have identified 'subskills' under the umbrella of 'comprehension skills' and there is a possibility that the findings could enlarge the list of comprehension skills and subskills into hundreds. In general, after reviewing various sources Rosenshine (1980, p. 540) observed that reading comprehension skills are commonly agreed to be:

...recognizing sequence, recognizing words in context,
identifying the main idea, decoding detail, drawing
inferences, recognizing cause and effect, and comparing
and contrasting.

Comparatively some of the seven skills from the consensus are not observable from the hypothesised hierarchy of subskills as postulated by Lunzer and Gardner's (1979) study. This is due to the fact that the eight subskills as postulated by Lunzer and Gardner's (1979) study are chosen to match the age-level of the subjects and purposely categorised according to the degree of difficulty of the questions asked. Such an order of categorisation of the eight subskills is assumed to be able to reveal the hypothesised distinct subskills. In this study the term 'subskills' refers to the postulated hierarchy of eight subskills as used in Lunzer and Gardner's (1979) study. The eight subskills and the corresponding definitions are taken from Lunzer and Gardner (1979, p. 44) and Waite (1980, pp. 21-22) are listed in Chapter 1.3.

1.7.4 Bahasa Melayu as the First Language (L1)

A precise historical review of the Malay language is not possible but for this study a general idea of the language merits a short explanation. According to Haji Omar (1992, pp. 155-157) Malay '...has been the lingua franca of insular Southeast Asia from time immemorial...the indigenous language of Peninsular Malaysia...[has] become the national and official language of three nations: Malaysia, Indonesia and Brunei. In Singapore, it is the national language...'.

Bahasa Melayu which according to Haji Omar (1992, p. 157) literally means 'the language of the Malays', is the language of instruction in primary, secondary and tertiary education in Malaysia. The term 'Bahasa Melayu' is used as the title of the latest primary school curriculum. Haji Omar (1992, p. 157) further says 'However, racial upheavals in the sixties, in which language was an important issue, had motivated the government of Malaysia to rename her national language *bahasa Malaysia*, viz. the language of Malaysia or the language of the Malaysian.'

What is important is the fact that the language and its various terms is still the same lingua franca learned in schools and spoken widely throughout the country.

In this study, the term is used in the context of the original English comprehension texts which were translated into Bahasa Melayu or, as it is commonly called, Bahasa. In the light of this study, especially in the context of the Bahasa syllabus, it is important to reflect on the specific reading skills as taken from the related syllabus. In the 'Syllabus content for the Malay language' (Ministry of Education, 1995, p. 7) reading skill is thought to consist of:

...pre-reading skill, the mechanics of reading and reading with understanding. Pupils should be able to read with fluency and correct pronunciation, intonation and rhythm and be able to understand and identify the main important idea. (Translated)

Clearly, the choice of the hierarchy of subskills and the testing of the subskills are in accordance with the syllabus prescribed. This is further supported from the related syllabus where reading and understanding in the Malay language are said to be comprised of:

Reading and understanding

- 2.17 Reading and understanding words and sentences.
- 2.18 Reading and understanding various types of literary and non literary materials.
- 2.19 Understanding the meaning of words according to context.
- 2.20 Stating the important information in the reading material.
- 2.21 Giving an appropriate title for a story or any genre.
- 2.22 Making a prediction on the conclusion of a story read.
- 2.23 Identifying the main character from a dialogue's material and acting out a script.
- 2.24 Locating and gathering main points from the reading material.
- 2.25 Reading and understanding symbols from mathematics and maps found in the reading materials.
- 2.26 Giving an opinion on any statement read.
- 2.27 Telling the lessons found from various reading sources.
- 2.28 Summarizing from reading material.
- 2.29 Making references in getting information.
- 2.30 Referring to dictionary in finding specific meanings, synonyms and antonyms.
- 2.31 Extensive reading.

Translated from: Kementerian Pendidikan Malaysia
(1995, pp. 8-9)

In view of the above syllabus content for the teaching and learning of reading comprehension, it is appropriate to test the assumed hierarchy of subskills in the identified age group. It is thought that the students, aged 12, are at the final year

of the primary schooling and should have learned all the reading skills as stated in the syllabus content.

1.7.5 English as a Second Language (L2)

The evolution of the status of English in Malaysian context merits a long historical review. This study does not intend to do so and in general reflects the contemporary development of English as one of the two main compulsory languages taught in the government primary and secondary schools and higher learning institutions. This status is describes by Haji Omar (1992):

...as stated in the Education Policy (1971) that English is the second most important language, second only to the national language. In terms of emphasis given to English in the schools and universities, and in terms of the exposure of Malaysians of all ethnic groups to English, as well as in terms of the function of English in social interaction especially in the urban areas, English is definitely a primary language but one that is second to Malay. (p. 114)

In relation to the scope of this study it is important to review the concept of an education system called The Integrated Secondary School Curriculum (ISSI) which was first initiated in 1980 by the Ministry of Education (Husin, 1992). ISSI in the context of English language for the secondary school stresses the integration of four basic skills (listening, speaking, reading and writing) which are geared to producing communicative competence in school leavers. In a broader context, the National Education Philosophy, which could be regarded as a student-centered policy, stresses the objective of developing students' spiritual, emotional, physical and intellectual growth. In this respect, the English Language Programme of the ISSI focus on the holistic development of the students and that it should produce effective communicators in English (Ministry of Education's Compendium, 1989).

The reading skills, which are viewed in the communicative paradigm, are illustrated in the syllabus content for form III as comprising a vast series of objectives that embrace a whole array of reading genres. (See Appendix A). What is important is that the eight categories of subskills as used by Lunzer and Gardner (1979) are directly and indirectly observed in Appendix A and Appendix B. This is especially true in deciphering Figure 4.2 (See Appendix B) where the concept of a hierarchy of learning difficulty is observable and thought to exist in the continuum of the processing strategies. As far as the readers' tasks is concerned it is assumed that 'graphic' is the easiest and 'discoursal' the most difficult to learn. Ideologically this view, where the processing strategies are thought to be in a linear fashion of increasing difficulty, correspond with the initial hypothesis of Lunzer and Gardner (1979) that comprehension comprises a multiplicity of distinct subskills. This initial assumption may be premature because the concept of a 'distinct hierarchy of processing strategies' is not directly stated in the compendium but is assumed to exist in a linear fashion.

1.8 Organisation of the Thesis.

The research reported in the following chapters is divided into two parts. In general, chapters 1-5 of Part I of this study document a small survey done from 1992 to 1994 on the issue of subskills in reading comprehension. It is an attempt to replicate the work by Lunzer and Gardner (1979) to find out whether reading comprehension comprises a set of discrete and hierarchic subskills in Malaysian schools. It is also an attempt to extend such questions to the world of learning English as a second language in schools. Lunzer and Gardner's (1979) initial hypothesis was that reading comprehension comprises a set of distinct subskills. This initial conviction was proven false and Lunzer and Gardner (1979, p. 64) concluded that differences in reading comprehension '...reflect only one general aptitude: this being the pupil's ability and willingness to reflect on whatever it is he is reading.'

In this survey the main hypothesis is set in a question: Does Lunzer and Gardner's hypothesis that reading comprehension is a unitary competence hold for readers in an L1 and L2 context in selected primary and secondary schools in Malaysia? This question led to a series of four separate comprehension tests on a total of 300 school children aged 12 and 150 secondary school students aged 15. The question types, which were specifically built to represent the hierarchy of the likely subskills, were arranged in order of difficulty. The data from the tests were carefully analysed using a series of analysing procedures such as finding the correlation matrix, Pearson-Correlation Coefficients, Principle-Component Analysis and Oblimin Rotation. In general it involved a robust analysis using the Factor Analysis procedures. The results of the study were encouraging and in general supported the findings of Lunzer and Gardner (1979).

Part II of this study, which began in the summer of 1994, comprises the remaining four chapters. In essence, they explore the comprehension strategies reported verbally by the students immediately after completing the comprehension tasks. In principle the immediate verbal protocols explore the reasoning processes or arguments used by the students for every written or chosen answers of the comprehension tests. A total of 16 students aged 15 were

selected from several secondary schools in the district of Johor Bahru, the capital state of Johor, Malaysia. Categorically, they represented the good and the average students in terms of overall academic achievements and in reading comprehension. They were required to answer the four comprehension tests which were equally divided into L1 and L2 and subsequently each of them had to verbalise the reasons for the answers they had chosen or written. From a total of 64 interviews only 16 were chosen to be analysed. Each interview was rigorously transcribed and translated to English language. Later, subsequent quantitative analyses were made to categorise the verbal protocol into the 8 categories of subskills. The findings of the analyses confirm the findings in the Part I Study and in general extend the understanding of the underlying processes of the reading comprehension strategies used by the students.

The main objective of the Part II study is to identify the relationship between the types of question which are based on the eight subskills and the coded discourse. In this respect, three general questions are asked. First, whether the good and the average comprehenders have remarkable degree of differences in answering the passages across the subskills. This is where a clearer understanding could be achieved as to the nature of interaction between the so called "lower order" items and "higher-order" items in the verbal discourse of the answered texts across the two groups. In other words, it specifically asked what subskills coloured the reasons for each test item and what supporting or compensating subskills come hand in hand with that item. The results could cast some light in understanding the mental processes involved in both groups and this could bridge more understanding on the nature of the processes and in assisting poor comprehenders to comprehend better. It is speculated that both groups have equal degree (or about the same) of similarity in using both the lower and higher order skills in responding to each test item. Thus, a variation of manners used by both groups in arriving at any answer or product is anticipated but the nature of the manners is not known.

At this juncture it is important to note the nature of prior learning experiences in shaping different ways of solving a test item (Bloom et al 1956. p. 16). In other words, each student in the two groups may choose different skills or strategies. Second, whether the good readers use multiple subskills more often than the average readers and finally whether the different languages produced different patterns of responses across the eight subskills. In this regard, the principal hypothesis is that the good readers use more subskills per test item than the average readers in all the four texts and in both languages.

It was hoped that the results would shed light as to understanding the nature of reading comprehension in L1 and L2 and thus allowing reading teachers and other interest groups such as curriculum planners and reading theorists to be better informed and equipped with relevant diagnostic tools which could be developed from this study in identifying problems in reading comprehension in primary and secondary schools in Malaysia.

The thesis ends with the discussion of the implications from both studies especially to the reading curriculum, instruction, pedagogy and classroom practice.

Chapter 2

Review of Literature Pertinent to the Part 1 Study

21. Models of the Reading Comprehension Processes in the First Language.

The following information discusses the underlying basic concepts of major reading models. The origin, the unique features and the debatable issues raised for each model are also reviewed.

2.1.1. The Bottom-Up Model

Gough (1976) proposed an influential reading model whereby the reader is thought to begin reading when the eyes first meet the letters on the page. The model is a complex explanation of the attempt to elaborate the information processing activities of a 'moderately skilled adult reader' during his first second of reading. Gough (1976) assumed that the initial eyes fixation is very rapid and this action leads to the formation of a visual pattern being reflected on the reader's retina. Gough (1976) used the term '*iconic representation*' to represent the rapid visual image and noted that the process of the formation of the iconic representation is extremely complex.

What is clear is that Gough's model highlighted the idea that reading is thought to be a rapid letter-by-letter identification, progressing to the meanings or to the higher levels of encoding the text. The input from each stage is recorded and passed on to the subsequent stage. It implies the mastery of specific discrete skills in sequence. In a similar tone, the graphic and visual inputs string from the time the eyes meet the text are sequentially transformed from the beginning of the graphical or visual inputs to the meanings of the text. A basic sequential processing pattern for young and old readers of the bottom-up model is best illustrated by Davies (1995, p. 58) as consisting of:

1. Eyes look
2. Letters identified and 'sounded out'
3. Words recognized
4. Words allocated to grammatical class and sentence structure
5. Sentences give meaning
6. Meaning leads to thinking

Gough's model is clearly a bottom-up process. It would seem that a reader's approach to comprehending a text is in a linear-fashion by first processing small chunks of information and then moving to larger bits of information until it becomes meaningful information. But, strictly speaking, Gough (1976) provides no clear explanation of the interaction of the sequence of events within the model itself, and the events have not given any clear clue to the use of other information or mechanisms outside of the text other than the use of guessing. Gough (1976, p. 532) argues:

A guess may be a good thing, for it may preserve the integrity of sentence comprehension. But rather than being a sign of normal reading, it indicate that the child did not decode the word in question rapidly enough to read normally. The good reader need not guess; the bad should not.

Gough assumed that the good reader has the ability to decode the information without making use of contextual information at all. According to Gough (1976, p.526):

The reader converts characters into systematic phonemes; the child must learn to do so. The reader knows the rules that relate one set of abstract entities to another; the child does not. The reader is a decoder; the child must become one.

Gough's model demonstrates the probable behaviour of a good reader in processing graphical input without relying on context. Guessing is thought to be a probable action taken by the reader when decoding fails to continue. However Nicholson (1993) found that about a decade ago Gough's view on the use of context in decoding words changed to some degree: Gough valued the use of a highly predictive context in facilitating word meaning but insisted that most words are not predictable and thus could only be read in a bottom-up manner. This view is in total contrast with the top-down model which gives more priority to making meaning from the text.

2.1.2 The Top-Down Models

In contrast to the bottom-up model, the top-down models in reading are driven by meaningful language units. Goodman (1970) rejected the sequential processes from learning letters identification to the larger language units. Goodman (1970) stressed that reading is 'a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the reader's expectation.' It is selective in the sense that the graphic cues are held for a short period in the *short-term memory*. The reader has to decide the likely meaning of the graphic cues or words before it is transferred to the *medium-term memory*. More selective processes continue and these lead to more possible meanings of the text which are again tested on whether the accumulated graphic cues fit the reader's expectations, syntactically and semantically. If there is a good fit between what is read and what was guessed earlier, meaning is achieved and this is stored in the *long-term memory*. If there is a mismatch between what is stored and what was guessed, the reader may regress to the related mismatch to identify the inconsistency that has occurred.

This whole model of information-processing is what Goodman (1970) called a psycholinguistic guessing game. In other words, the reader makes hypotheses and uses the language cues in making intelligent predictions and expectations of the meaning of the material read. Smith (1973), another advocate of the meaning-based approach to reading, criticised those who support the bottom-up model on

the teaching of phonics skills on the ground that '...any attempt to read by translating letters to sounds through the application and integration of phonic rules could result only in catastrophic overloading of short-term memory.'

In the top-down models (Goodman, 1970; Smith, 1973) the reader is thought to have certain reading tasks and this carries the interpretation that reading comprehension is an act of hypothesis testing. The reader, who is guided by the selective and predictive processes, concurrently integrates three kinds of information known as graphic, syntactic and semantic in making sense of the text. The reader makes use of minimal information from the text and relies more on the context of the text to confirm his predictions and anticipations of the text read (Goodman, 1970). In a similar vein, Smith (1973, p. 188) says 'Fluent readers do not read words, they read for meanings...in order to read one must constantly form expectations that reduce the uncertainty of what one is reading, and therefore reduce the amount of visual information required to extract its meaning.' Still, Goodman failed to provide a detailed account of the integrative process of the syntactic, semantic and the phonological inputs.

Goodman (1976, p. 498) thought that 'Efficient reading does not result from precise perception and identification of all elements, but from skill in selecting the fewest, most productive cues necessary to produce guesses which are right the first time.' This idea is refuted by Stanovich (1986) on the ground that current research using eye movement technology have proven that fluent readers do not behave as Goodman thought. Stanovich (1986, p. 368) says 'The good reader is not less reliant on the visual information, but the good reader does allocate less capacity to process this information.' This is further supported by research (See, Bernhardt, 1991, pp. 73-74) in the eye-movement protocol of a fluent native (college-level) and a fluent non native reader of English: The native reader of English fixates or "sees" about 84% of all the content words in a passage titled "Acid Rain" and about 17% of the function words. The fluent non-native reader of English, a Japanese professor, fixates even more content words on a much simpler text, titled "Travel Log", than the native reader. In relation to word recognition it is clear that skilled or fluent readers do give more attention to words than Smith (1973) thought.

Clearly, Goodman and Smith stressed the importance of "top" knowledge the readers bring to the page. Albeit the criticism on Goodman and Smith, Willinsky (1990) found that the model is proven to exist in understanding children's literacy. Willinsky (1990, p. 69) says '...the novice reader proceeds to work *down* from this general level of expectation to a reading of the individual words with increasing accuracy, based primarily on expectations of what would follow in the story of the sentence.' Davies (1995, p. 58) illustrates the top-down model as a processing sequence comprised of:

1. Eyes look
2. Thinking-predictions about meaning
3. Sample sentence as a whole to check meaning
4. To check further, look at words
5. If still uncertain study letters
6. Back to meaning predictions.

Both models begin with different premises. What is notable is the fact that meaning is given early emphasis in the top-down models where the reader is active in making predictions and sampling parts of the reading material. In contrast the bottom-up models view reading primarily as a decoding process: the reader is thought to reconstruct meaning by systematically working on the smaller textual units of language such as the letters and words and moving to the largest units of the language. Still, Beard (1990) observed that despite the views of Goodman (1970) and Smith (1973) that reading emphasises getting the meaning from the text, both of them are eclectic in their suggestions for the teaching of reading: Goodman and Smith are not rejecting the sensible use of phonics instruction in meaningful ways. Perhaps, meaningful phonics instruction was overlooked by Turner (1990) when he criticised Smith's and Goodman's models of reading processes such as the 'guessing game' as being the causes of the decline in the standard of reading in England. Turner's criticisms (1990, p. 7), which are thought not to be critical enough by some reading scholars, came from the belief that the 1980s "progressive" ideologies enshrined in various reading programmes such as 'real books', 'story method', 'emergent reading', 'wholistic approach', 'osmosis' and 'apprenticeship approach' were '...associated with failing

school children at class level, school level, local authority level and, thus far, regional level (the south and south-east) in Britain.' In general, Turner's criticisms are not fine-tuned or lack scholarly in-depth debate on the detail understanding of the so called 'psycholinguistic guessing game' of Goodman (1976) or the other similar schools of thought that emphasised meaning in reading.

Harrison (1996) summarised and highlighted the fact that scholarly critics made on the 'psycholinguistic guessing game' of Goodman (1976) are mainly aimed on the issue of how the good readers read. Four main issues of the critics as noted by Harrison (1996, p. 11) are:

- Goodman's model is poor on detail.
- Good readers are not dependent on context for word recognition.
- Good readers fixate nearly every word as they read.
- Good readers have automatic word recognition.

Harrison's review (1996) of the above scholarly criticisms in reading process is important to classroom practices. According to Harrison (1996, p. 17) it is important that in order to produce proficient readers, reading teachers need to consider some implications of the above criticisms on the following points:

- automatic [and] rapid letter recognition.
- rapid word recognition.
- the ability to use context as an aid to comprehension [and]
- the ability to use context when necessary as a conscious aid to word recognition.

Commenting further on the above four points, Harrison (1996, p. 17) emphasised that in reading pedagogy, reading teachers should be aware that the four implications should not overlook the need of meaningful reading.

2.1.3 The Interactive Model

This model is widely regarded as the current reading model by reading psychologists (Harrison, 1992). The model is best represented by the work of Rumelhart in the mid 1970s. Davies (1995, p. 63) claims '... [Rumelhart's interactive model] is currently the most influential model underpinning both L1 and L2 approaches to reading, ... '.

In principle this model assumes an interaction of information drawn from the readers' knowledge (top-down) and the text (bottom-up). Thus, it is a complex process of the perceptive and the cognitive skills of the reader. However, Rayner and Pollatsek (1989, pp. 467-468) criticised Rumelhart's model in that it has no explanation on how the various sources of information interact, particularly:

...about the basis on which various kinds of hypotheses are generated nor does it specify the relative importance of the contribution of the various knowledge sources (syntactic, semantic, orthographic, and lexical)...., control of eye movements, the phonological route in word recognition, or comprehension issues beyond the level of the sentence.

The model is thought to emphasise the interactive nature of the parallel processing mechanisms which are observed to be practised by fluent skilled readers. Davies (1995, p. 64) says '... [in Rumelhart's model] the reader is seen to be able to draw simultaneously, but selectively, upon a range of sources of information, visual, orthographic, lexical, semantic, syntactic and schematic.'

In principle, the interactive model is a balanced information processing concept that does not strictly adhere to either the bottom-up or the top-down models. It projects a view that information processing can be drawn from various knowledge sources and that it is simultaneous and selective in behaviour. Understanding is synthesised from the various knowledge sources.

2.1.4 Rayner and Pollatsek's Bottom-Up Interactive Model

Rayner and Pollatsek's (1989, p. 472) reading model is '...primarily a bottom-up model, but top-down processes do interact with the bottom-up processes...[and] ours will be somewhat vague about the role of higher-order processes.' The model is based on the observation of fluent adult readers. The process of encoding sequence is thought to begin when the eye fixates on the page and this is drawn heavily from eye movement studies. The encoding actions of the eyes are thought to be simultaneous and yet separate processes.

The first encoding is known as foveal word processing and is mainly concerned with the letters of the words in focus. The second encoding, known as the parafoveal processing, is the perceptual span of the eyes in extracting information to about 15 character spaces to the right side of the initial fixation of the eyes. These initial encoding processes are subsequently followed by the lexical access process. This lexical process may work with the parafoveal process in a rapid manner. Lexical meaning can be derived through a direct route (vocabulary stored in the Long-Term Memory) or an indirect route. In the latter case, the reader may apply rules (grapho-phonetic) or analogies or the combination of both rules and analogies in creating an auditory code/inner speech, where recent information read is held temporarily in the working memory. The indirect route is thought to be automatic and Rayner and Pollatsek (1989, p. 474) say that it '...probably serves as a subsidiary system for lexical access of familiar words.'

The Rayner and Pollatsek's (1989) model exhibits the recent development in the eyes movement studies and stressed the importance of 'bottom-up' without ignoring other cognitive inputs in processing the text such as background knowledge, syntax and semantic clues. Rapid or automatic word or visual recognition would therefore allow more time for comprehending the text. In relation to L2, Davies (1995) observed that such rapid word recognition processes, as conceived by Rayner and Pollatsek, is important in L2 or else L2 beginner readers have to take more time in processing the visual inputs.

2.1.5 Stanovich's Interactive-Compensatory Model

Stanovich's compensatory assumption is based on the hypothesis that a 'balancing act' takes place when a reader encounters a problem in understanding a text. In other words, when the reader is poor or deficient in any knowledge source, he or she may rely more on other knowledge sources. This is to say that the interactive-compensatory process assumes that a reader may make use of the various knowledge sources available in a simultaneous and interactive fashion at any level of the processing hierarchy (Stanovich, 1980).

Stanovich (1980) invalidates the bottom-up models on the basis that higher-level reading processes do not necessarily rely on the completion of the lower-level processes. Equally, the hypotheses of the top-down models are challenged on the ground that 'The finding that in some situations poor readers rely more on context than do good readers presents problems for top-down models, which hypothesize that reading becomes more conceptually driven as fluency develops.' (Stanovich, 1980, p. 47).

Stanovich assumes that poorer readers may rely more on higher-level knowledge sources to compensate the problems encountered in the lower-level knowledge sources such as word recognition. In contrast, good or fluent readers' word recognition is automatic and this rapid word recognition suggests that the good readers may have more time to interpret the text than the poor readers. Harrison (1992) says that Stanovich's view of automatic word recognition is in sharp contrast with that of Smith (1971) and Goodman (1970). Harrison (1992, p. 11) says:

Goodman is now thought to have been wrong in suggesting that in fluent reading only minimal text cues are sampled in the word recognition part of the reading process...Smith is thought to have been wrong in suggesting that word recognition is 'a hindrance' to fluent reading.

Reflecting on the behaviour of reading subskills in an interactive-compensatory manner, Stanovich (1980, p. 15) says:

...an interactive model, when coupled with the assumption that various component subskills of reading can operate in a compensatory manner, leads to reconceptualization of the nature of individual differences in reading.

This implies that in the quest for meaning a deficit in any subskills in reading comprehension may be compensated by other subskills. It appears that the interactive-compensatory model embraces a holistic view of reading and neither the top-down nor the bottom-up models exhibit such a compensatory element. Stanovich (1980) suggested that readers who are poor in context-free word-recognition may consciously use the context in identifying the meaning of the word and this action '...leaves fewer cognitive resources left over for comprehension operations that work on integrating larger text units.' On the other hand, other than the rapid and automatic context-free word recognition by visual or by phonological recording, good readers' performances, to some degree, are thought to be better than the poor readers in '...strategies for comprehending and remembering large units of texts...[allowing] more attentional capacity...for integrative comprehension processes.' (Stanovich, 1980, p. 64).

Generally speaking, it is sensible enough to accept Goodman's idea that good readers are efficient in using context in the quest of meaning but Stanovich (1986, p. 370) observed that '...reading skill is not determined by skill at contextual prediction, but rather that the level of word-recognition skill determines the extent to which contextual information will be relied on to complete the process of lexical access.' This gives the idea that context is useful to poor readers in that it provides clues in finding the meaning or identity of a word that hinders understanding. On the other hand, efficient automatic graphemic word recognition displayed by good readers means that there is no need to play a 'guessing-game' and thus more time could be spent on comprehension. Nicholson (1993) observed that Gough, the proponent of the

bottom-up model, recognised the Stanovich model in using guessing by good readers when decoding fails to operate.

2.2 Models of the Reading Comprehension Processes in the Second Language

Contemporary views of second language reading have been influenced by first language research findings. Bernhardt (in Hulstijn and Matter, 1991) observed that in the past decade the best characteristics of second language reading research were framed in two general questions: whether second language reading is a language proficiency problem or a reading problem. Later, a third question was added: whether reading in a second language is a knowledge problem. Bernhardt thought that the term "problem", if thought as a 'disability' as suggested by Alderson (1984), may not be appropriate in understanding second language reading process. Rather, second language reading, now thought as 'a complex social and psycholinguistic process that cannot be separated into reading components and language components' should be viewed as 'a new and different *literacy* (Bernhardt, in Hulstijn and Matter, 1991, pp. 31-44).

In the past, second language reading was conceptualised as the bottom-up model as in L1 and with this perspective Carrell (1987, p. 24) says 'Problems of second language reading and reading comprehension were viewed as decoding problems.' In the mid 1970s the psycholinguistic top-down models began to make an impact on the second language reading model. This is observed by Bernhardt (1991) and Grabe (1991) who found that models of reading comprehension processes in second language reading are overwhelmingly dominated by the Goodman (1970) and Smith (1973) psycholinguistic reading perspectives. Although it is not a dominant model in L1 reading research, Bernhardt (1991, p. 22) thought that its existing dominance in L2 reading research may be due to the fact:

...that academicians in this field have agreed that
the psycholinguistic framework provides the most
viable explanation of reading a second language...

[and] there is a basic lack of awareness and perception of the capabilities of models other than those of Goodman (1968) and Smith (1971) to explain second language reading phenomena.

From the mid 1970s to the present situation, as Grabe (1991) notes, reading in ESL has been very much influenced by the work of Clarke and Silberstein (1977) and Coady (1979). These authors have adapted the psycholinguistic models to the needs of the second language learners and '...many of these instructional implications still remain as important guidelines though no longer motivated by the psycholinguistic model explanation.' (Grabe, 1991, p. 377). Bernhardt (1991) observed that there are insufficient studies in second language reading research and that about a decade ago a little over a hundred of such studies had been conducted. This is extremely small when compared with thousands of such research projects in L1. In view of the problem, Grabe (1991, p. 378) says 'A primary goal for ESL reading theory and instruction is to understand what fluent L1 readers do, then decide how best to move ESL students in that developmental direction.'

Bernhardt (1991) attempts to provide an initial holistic view of second language reading merits a discussion. The model is called 'A Reader-Based Perspective' and is characterized by the interaction of texts' linguistic elements, structure, pragmatic nature, intentionality, content and topic with the reader's objectives of reading a text. A reader of a text may try to make sense of it by deciding what is important or may "reconstruct" the important ideas or points in accordance to the decisions. In other words, as Bernhardt (1991, p. 15) says '...certain readers "see" different things in texts.'

2.3 Two Contrasting Theories in Reading Comprehension and Their Implications.

It has been suggested that in any research study the importance of explanatory and consistent theories in steering any effective research is crucial in producing a good research (Kamil, 1984, pp. 43-44). This is especially true in highlighting the main theoretical arguments in the area of reading research. In relation to Kamil's suggestion, this study investigates one of the theoretical problems in reading comprehension research through two common and related issues: whether there is such a thing as distinct reading comprehension skills or subskills and whether the skills or subskills are in any hierarchical cumulative order (Lunzer and Gardner, 1979; Rosenshine, 1980). This theoretical issue sets the guiding framework of discussions and debates that have risen from many related studies.

To begin with, the historical origin of the debatable theoretical issue need to be briefly addressed. Samuels (1980) claimed that the debate between the holistic-meaning-emphasis, the subskills and the mixed approaches in teaching reading is not new but has a long recorded history of more than two thousands years. The debate on whether to teach in the manner of the Greek alphabetic ABC method or the so called "natural" whole-word approach stormed Europe and America in the mid 1800s. Still unresolved, the arguments dragged on until 1870 when the whole-word method started to be accepted by educators. By the mid 1950s the phonic method was once more advocated with the publication of *Why Johnny Can't Read and What You Can Do About It* by Rudolf Flesch. Still, Samuels (1980, p. 204) noted that both views are not without drawbacks:

A major criticism of the [Greek ABC] method was that spelling the word before pronouncing it interfered with comprehension which led some educators to advocate a different approach...Flesch argued that children who were taught by the whole-word method had difficulty because of their failure to acquire word analysis skills.

Many psychologists refer to reading as "a skill" which involve many complex processes, including cognitive as well as attitudinal and manipulative. But this all-embracing term, better known as a holistic view used by the psychologists is thought to be in conflict with another definition of "skills" used in the field of education particularly the reading instruction materials of the 1970s and 1980s. The mastery of "skills" before one can become a successful reader is exemplified as the teaching and learning of separate subskills that begins with word discrimination and then proceeds to vocabulary development and finally comprehension. This approach is heavily criticised particularly by Goodman (1982) and Downing (1982). Goodman (1982, p. 48) argues that 'No researcher has ever been able to support any particular sequence of skill instruction as having any intrinsic merit which derives from linguistic and psycholinguistic analysis.' The danger of dividing the teaching of reading into 'code-emphasis stage' and then proceeding to 'meaning-emphasis stage' could also lead to ineffective comprehension. Reading should be taught as a single process that focuses on comprehension rather than concentrating on observed behaviour based on performance in tests (Goodman, 1982).

Numerous scholars in the relevant field have probed the nature of reading comprehension as unitary or an accumulation of a hierarchy of subskills. Their findings can be broadly divided into two groups; subskills and holistic.

2.3.1 The Subskills Reading Theory

In reading research, finding a set of subskills which are thought to comprise reading comprehension is not without theoretical issues. If such skills are identifiable it certainly would provide a systematic framework for the effective assessment and improvement of reading comprehension in schools. If we insist in finding such a set a theoretical framework should be laid out and analysed in the light of related research. The issues in reading comprehension skills have been known to be evolving around whether there are discrete reading comprehension skills and whether such skills are hierarchic in nature (Rosenshine, 1980). In order to understand the theoretical issues it is pertinent to

reflect on the previous major studies that have set the path of this study.

The historical development of the subskills proponents began as early as 1944 with the work of Davis. He found nine distinguishable skills in reading. Thorstone (1946), after reanalysing the 1944 data from Davis, said that Davis's 1944 conclusions from the complex correlations components should represent one common factor thought to be general reading ability. Twenty four years later Davis made another conclusion claiming that the nine skills could be reduced to four skills, namely; to identify word meaning, draw inferences, identify techniques used by a writer in creating the mood of the passage and find answers to questions (Davis, 1968). Thorndike (1971) re-factorised Davis's 1968 data and identified only 'word meaning' (Skill 1) as distinguishable as a separate skill and the other seven skills as not distinguishable. A year later, Spearitt refactorized Davis' data and identified four separate skills namely Skill 1, Skill 2, Skills 3 and 4 and Skill 5 (Spearitt, 1972).

Davis (1968) found that the reading comprehension of the tested 988 twelfth-grade pupils was not unitary but involved subskills. The tested skills were:

1. recalling word meanings
2. drawing inferences about the meaning of a word from the context.
3. finding answers to questions answered explicitly or in paraphrase.
4. weaving together ideas in the content.
5. drawing inferences from the content.
6. recognizing a writer's purpose, attitude, tone and mood.
7. identifying a writer's techniques.
8. following the structure of a passage.

Then, according to Spearitt (1972) in 1971, Thorndike refactorized Davis' data and suggested that except for word knowledge, the reading skills were not separately distinguishable. Spearitt (1972) analysed the data by subjecting them

to a uniqueness analysis and a varimax-rotated principal components analysis. Then he refactorized Davis's data using the maximum likelihood factor analysis. Spearitt (1972) concluded that out of the eight subskills postulated by Davis (1968, p. 109) only four skills known as '...recalling word meanings; drawing inferences from the content; recognising a writer's purpose, attitude and tone; and following the structure of a passage...' are distinguishable skills, thus confirming the findings by Davis (1968) except in Skill 3, finding answers to questions answered explicitly or in paraphrase, which Spearitt (1972) failed to prove as a separate skill.

Spearitt's (1972, p. 110) analysis acknowledged the fact that 'Vocabulary is the best differentiated, as in both Davis and Thorndike analyses.' What is notable from Davis's (1968) findings is the perception of reading as a process composed of distinct subprocesses. These findings strengthened the assumption that comprehension can be analysed into various discrete hierarchical subprocesses which are necessary for successful performance in producing good readers.

To simplify the above matter an outline of the historical development of the correlational studies in identifying the distinctiveness of different comprehension skills in reading is drawn from Spearitt (1972, p. 94) . They are:

<u>Years</u>	<u>Quoted remarks by Spearitt (1972, pp. 94-98)</u>
1917--	Thorndike '...concluded that reading comprehension was basically a process of reasoning.'
1944--	Davis '...concluded that there were nine distinguishable skills in reading.'
1946--	Thurstone challenged Davis' findings and stressed that '...the correlations could be adequately explained in terms of one common factor, presumably representing differences in a general reading ability.'
1968--	Davis reviewed the experimental studies of reading comprehension and '...selected eight skills and constructed items to measure each of these skills.' One item referred to

one passage '...in order to avoid the problem of experimental dependence among responses to items.' The items were tested on 12th Grade High School Students. He used a uniqueness-analysis technique to analyze his data. Davis found that '...reading comprehension among mature readers is not a unitary trait, and that substantial parts of the mental abilities used in the eight skills are independent of one another...Skills 1,3,5,6 and 8 are experimentally distinguishable.'

1971-- Davis used the 1968 data using principal components analysis after rotation to a normalized varimax criterion.

1971-- Thorndike used Davis' 1968 data using components analysis and '...showed that the non-chance variance in the sets of eight tests could be completely accounted for by three factors...His rotated factor loadings...showed that the word knowledge test(Skill 1) could be distinguished from the other reading tests in terms of its factor pattern, but none of the other skills was separately distinguishable...which he believes to be largely a measure of reasoning.'

What is clear is the fact that the same data produced different results and conclusions when analysed by the different methods of analysis. This is true because on the one hand the data was first used by Davis (1968) with his analysing technique known as Uniqueness Analysis and subsequently with Principal Components Analysis. On the other hand, Thorndike (1971) used Factor Analysis which is further extended by Spearitt (1972). Using the maximum likelihood factor analyses, Spearitt (1972) reanalysed Davis (1968) data and observed the existence of the four skills as discussed above.

Clearly reviewing the historical development of the subskills proponents signifies inconclusive findings and this is due to the different data analysing techniques. What can be observed is that the different methods in analysing the same original data from Davis (1968) yielded different skills and (Rosenshine, 1980) observed that '*remembering word meaning*' emerged consistently across

the analyses of the same data by Davis (1968), Davis (1972) and Spearitt (1972).

The proponents of the subskills approach favoured the concept of part-to-whole in mastering reading especially in early reading. They viewed reading as '...a complex phenomenon, consisting of various subskills. Up to 100 or more subcomponents have been defined and, based on component analyses, three to eight necessary and sufficient subskills have been identified (cf. Rost, 1985).' (Rost, 1989; p. 89). Farr (1969, pp. 2-3) found that many of the research attempts, as far back as in the 1940s, in finding the skills underlying reading ability were in fact oriented to factor-analysis in approach. Farr (1969) also stated that many of factor-analysis findings agreed on limited number of factors that constitute reading skills and due to the nature of the various studies direct consistency of the factors in all of the studies could not be established.

The subskills approach is dismissively viewed as traditional task-analytic behaviour and the related curriculum as an 'assembly-line' model. Dole et al.(1991,p.241) describes this 'island' view as follows:

..., novice readers acquire a set of hierarchical ordered subskills that sequentially build toward comprehension ability. Once the skills have been mastered, readers are viewed as experts who comprehend what they read. In this view, readers are passive recipients of information in the text. Meaning resides in the text itself, and the goal of the reader is to reproduce that meaning.

The above quotation is an indirect reflection on the effects of the bottom-up model in reading. However, this 'assembly-line' view of reading comprehension is declining considerably, particularly with numerous recent studies that place more emphasis on reading as an interactive process between the reader and the text (Dole et al. 1991). Reviewing the current syllabus content of English language for the primary school in Malaysia reveals a close resemblance to the 'hierarchy of subskills' paradigm in practice (Kementerian Pendidikan Malaysia, 1982). The hierarchical view is also adopted to some degree in the current reading curriculum of the National Curriculum for England and Wales despite '...signs in

some early documents produced in relation to assessment in English in England and Wales that it was recognised that variations in text difficulty posed problems for hierarchical models of reading comprehension.' (Harrison, 1994, p. 22). This is strongly supported by the famous reading comprehension research study of Davis (1972) who believed that his study had not proven the existence of hierarchy of comprehension skills. Davis (1972, p. 172) says:

The hierarchical skills theory cannot be reconciled with experimental findings concerning the intercorrelations of skills tests in reading comprehension...scores of 988 twelfth-grade students in academic high schools show no marked evidence the eight skills test can be arranged in a clear-cut order of cumulative agglomeration of simple skills in more complex skills. More systematic investigation of this point needs to be made.

2.3.2 The Holistic Reading Theory

Reading comprehension is currently viewed as an active meaning-making process. It is a cognitively based view that stresses the interactive and constructive aspects of reading, and not the traditional view that emphasises mastering many subskills which are thought to be mastered through drilling and practice and later on automatically applied to any text read. The difference between the old and the new views, as Dole et al. (1991) put it, is that the latter acknowledges '...the knowledge that students bring to the task and the strategies that they use to foster and maintain understanding...' (p. 241). Dole et al. (1991) also draw some important teaching implications out of the cognitively based view of reading comprehension. In essence, the cognitively based view propagates a holistic view of reading. It emphasises the conscious control of the readers on a set of comprehension strategies that can be effectively learned and taught. Dole et al. (1991, 243-249) summarise the comprehension strategies drawn from previous research into five headings known as:

- a) Determining importance
- b) Summarising information
- c) Drawing inferences
- d) Generating questions
- e) Monitoring comprehension

Harrison (1994) points out that there are a few close resemblances between the eight comprehension subskills as hypothesised by Lunzer and Gardner (1979) and the list of comprehension strategies highlighted by Dole et al. (1991). As an example '*Determining importance*' in Dole et al. (1991) closely resembles '*Finding salients or main ideas*' in Lunzer and Gardner (1979). But a crucial difference between the 'subskills' and the 'strategies' paradigms lies in the fact that the former consists of readers' internal unobservable automated processes and the latter are observable active on-going processes in making meaning out of the text read (Harrison, 1994).

According to Rost (1989, p. 89) the holistic view in reading comprehension is supported by research conducted by many scholars, namely; "... (Thorndike, 1917 a,b,c; Thurstone, 1946; Harris, 1948; Alshan, 1964; Thorndike, 1973/4; Drahozal and Hanna, 1978; Andrich and Godfrey, 1978/1979; Lyons, 1974)." Johnston (1983) acknowledged the fact that Thorndike, the main proponent of the holistic approach, stressed the importance of reasoning in teaching reading and not the separate subskills.

Johnston (1983, p. 4) says '...the main proponent of the holistic approach has been Thorndike (1974, p. 57) who claims "The barrier....[is]....not primarily a deficit of one or more specific and readily teachable reading skills but is a reflection of generally meager intellectual processes."' '

In the United Kingdom, Lunzer and Gardner (1979) found no conclusive supporting facts that reading comprehension consists of distinct subskills and '...no support for the hypothesis that some pupils might 'possess' lower-order skills but not higher-order skills...' (p. 300) and in general they concluded that "...individual differences in reading comprehension should not be thought of in

terms of a multiplicity of specialized aptitudes...[they only] reflect only one general aptitude: this being the pupil's ability and willingness to reflect on whatever it is he is reading." (p. 64). Subsequently, the famous Directed Activity Related to Text's (DART) project of Lunzer et al. (1984, p. 13) is based on the holistic premise and this is clearly stated as follows:

There is little evidence to justify the teaching of reading comprehension as a set of separate sub-skills....
These (the subskills) are not different processes in the reader. They are simply different tasks required of the reader to 'prove' that s/he has understood.

In another study, Chapman (1987) reviewed another Schools Council project based at Manchester University where a team headed by Southgate made their findings available to the public two years after the Lunzer et al. (1984) project at the University of Nottingham. Chapman (1987, p. 53) summarized Southgate's findings on the 1980s' trend of reading being defined as a thinking process and acknowledged Lunzer and Gardner's findings (1979), especially on reading comprehension not as a continuum of sub-skills but as a 'global act'.

In Germany, Rost (1989), tested the reading comprehension of 220 German Second Graders (114 girls, 106 boys), and found that reading comprehension is a holistic process related to general intelligence and verbal solving ability. His finding supports Thorndike's (1917) holistic approach.

Another prominent supporter of the holistic approach is Goodman (1970, p. 25) who claims 'There is no possible sequencing of skills in reading instruction since all systems must be used independently in the reading process even in the first attempts at learning to read.' Wallace (1992) deduces that the approaches of Smith (1971) and Goodman (1982) to reading are unitary in nature.

In summary, the above literature review proves one main fact: the findings do not support the existence of a hierarchical sequence of skills in reading comprehension. This is affirmed by Wallace's (1992, p. 42) statement that '...it is

not possible to identify specific skills which can be built up in any hierarchical way to produce an effective reader.'

In principle, regardless of the polarising positions taken up by reading scholars on the issue of the existence of reading subskills, it is necessary that a study is replicated and extended in other languages particularly in exploring verbal reasoning processes in reading comprehension in a second language. Such verbal reasoning processes are worth investigation and are fruitful in extending our understanding of the complexities of reading comprehension processes.

CHAPTER 3

Design and Methodology of the Part I Study

3.1 Overview of the Part I Study

This study investigates the process of reading comprehension in L1 and L2. The basis of the inquiry is whether reading comprehension skills can be built up in any hierarchical order in producing an effective reader and whether these comprehension skills are distinct in nature. In doing so, this study replicates a landmark study in the United Kingdom. A team headed by Professor E. A. Lunzer and W. Keith Gardner (1979), through the Schools Council project, directed a detailed study which is best stated by Lunzer and Gardner (1979, p. 39) as '...chiefly to inquire whether the several tasks which may be exemplified in comprehension-test items derive from distinct skills or subskills or whether comprehension is a unitary ability.'

At the beginning of the study the team was very much inclined to the idea that reading comprehension subskills were hierarchical in nature. Prior to their famous findings, Lunzer and Gardner (1979, p.40) acknowledged that their '...weight of opinion inclined to the first hypothesis, that comprehension involves a multiplicity of subskills,...'. This view is noted by Lunzer and Gardner (1979, p.41) as the widespread, influential and accepted view among educationists and the view was amplified by the publication of reading comprehension subskills such as *Reading in the Subject Areas, Grades 7-8-9* (1964) published by the New York City Board of Education (See Lunzer and Gardner (1979, p. 42). Lunzer and Gardner also acknowledged that many research attempts to identify discrete subskills in reading comprehension are far from conclusive. The study by Davis (1968) who carried out two 96-item tests on 98 college students and found four separable skills which he called 'identifying word meaning', 'drawing inferences', 'identifying a writer's technique in creating the mood of the passage' and 'finding answers to questions', had been influential in motivating the Lunzer and Gardner (1979) study. Davis's data and findings were later scrutinised by other researchers including Spearitt (1972) who acknowledged Davis's findings

but with a little modification to the fourth factor which he thought ability to follow the structure of a passage'. A review of the related literature is discussed in chapter 2.

The various interpretations of Davis's data to some degree influenced Lunzer and Gardner (1979) to carry out their landmark study on the nature of reading comprehension and in general they found that reading comprehension cannot be broken down into distinct subskills and is far from being in a hierarchical order. Lunzer and Gardner (1979, p.64) says:

We conclude that individual differences in reading comprehension should not be thought of in terms of a multiplicity of specialized attitudes. To all intents and purposes such differences reflect only one general aptitude: this being the pupil's ability and willingness to reflect on whatever it is he is reading.

This study is concerned with the debatable issues of the hierarchical or unitary skills in reading comprehension in L1 and L2. At the time of the writing up of this study it is observed that there is no similar or comparable study being done in Malaysian primary and secondary schools particularly in testing reading comprehension with extensive texts. Added to that is the importance of understanding the issues in two different languages and this is manifested by concurrently testing whether reading comprehension in the two languages, L1 and L2, is unitary or atomistic in nature. Thus, this study set out to replicate and extend Lunzer and Gardner's (1979) study in Malaysian primary and secondary schools. Why replicate the Lunzer and Gardner's (1979) study? Up to the writing of this study it is noted that no identical study had been carried out extending our understanding of the study either in UK or in Malaysia. Other than replicating the Lunzer and Gardner (1979) study, this basic research seeks to extend the testing of the reading comprehension tests in L2. It also seeks to evaluate the inferred assumption that hierarchy in reading comprehension is unproblematic as noted in the language syllabuses of the primary and secondary schools in Malaysia (Kementerian Pendidikan Malaysia 1982;1988;1989;1990).

3.1.1 Choosing the Research Method.

The research methodology of this study is commonly known as a descriptive method. The nature of this research is concerned with the existing conditions or the relationship of the reading processes of the samples studied. To be more specific, this study is called a small scale statistical survey study (Cohen and Manion, 1989; Brown, 1988). Tuckman (1978, pp. 10-12) suggested that statistical research should be systematic, logical, tangible, replicable and reductive in principles. In view of the above suggested principles, a systematic planning of the survey has adopted the suggested well-defined stages by Cohen and Manion (1989, pp. 97-123). This study has followed the suggested principles and the stages to a considerable degree. In other words, not all of the suggested stages are strictly practised in the study.

3.1.2 Chronology of the Study.

The following time frame details the chronology of events of the study:-

- a) November 1992-May 1993.
 - i) Identifying relevant objectives of the enquiry. The central aim was to identify the postulated subskills in reading through a reading comprehension test.
 - ii) Replicating the work of Lunzer, Waite and Dolan (1979). With the help of the supervisor a letter of approval, dated 15/2/93, was granted from Maurice Waite. He allowed the use of the four comprehension tests.
 - iii) The four original comprehension texts were translated to L1. The contents and the flow of the texts were well retained. Only a few modifications were made to the original texts. This was to suit the readers' socio-cultural backgrounds.
 - iv) The four texts were then checked for any discrepancy that could have occurred during translation. This was done by a senior linguist in Bahasa Malaysia (L1) from the Department

- of Education, International Islamic University, Malaysia.
- v) The eight texts, four in L1 and four in L2, were randomly selected and tested on four language teachers from Malaysia. They were reading for their Bachelor of Education in Teaching of English As a Second Language at the University of Nottingham. The teachers had experience of teaching reading comprehension in Malaysian primary schools for many years. Later, they were interviewed and in principle they commented that the tests were relevant in terms of their readability and the degree of difficulty in the comprehension questions.
 - vi) Three Malaysian teenagers, aged 13, 14 and 15 were randomly selected and tested on all the texts. They were from the Nottingham area. Their suggestions and comments were taken into consideration. As an example one of them commented on the length of the texts and suggested that a shorter version for all the texts would be better rather than the lengthy ones. Due to the nature of this replication study, the suggestion was not applicable.
- b) June 1993- August 1993.
- i) A letter of approval from the Educational Planning and Research Department, Kementerian Pendidikan Malaysia, to conduct the study in any state in Malaysia.
 - ii) A letter of approval from the Education Department of Johor to conduct the study at the selected schools.
 - iii) Meetings with the headteachers and the language teachers of each school. In principle, they agreed that the tests were appropriate for the students with a minor alteration to the prior planning. At first, it was planned that the testing in both languages were to be administered to all the samples. Then, suggestions from the headteacher and the language teachers from one of the primary schools had to be taken

into consideration. They expressed the feeling that the L2 texts were difficult and too long for many of the pupils to cope with. After a lengthy discussion in terms of time limitation, the classroom schedules and other relevant factors, we compromised to resort to the L1 texts. It was also observed that the samples were to sit for the national Primary School Assessment Test in the months of September and October of that year. This implies that the samples were at the end of their learning of the primary school curriculum. A total of 300 students from the two schools were tested on the four L1 texts. Only the responses from 271 students were valid for the Factor Analysis.

- iv) Considering the time factor faced by the affected secondary school teachers, another compromise was met. This time only the L2 texts were permitted to be tested on the samples. The headteachers and the affected teachers believed that the L1 texts might not be challenging enough but somehow agreed that the degree of the difficulty level of the L2 texts matched the targeted students. A total of 150 students from the two schools were tested on the four L2 texts. These students were to sit the Lower Secondary School Assessment Test in October. Thus, it could be judged that the language curriculum of the lower secondary school should have been delivered to the students. Assessing the written answers yielded valid responses from only 138 students. 12 students had failed to sit all the four texts. They were identified to be absent on the testing periods on some of the tests or were on other unavoidable school activities.

c) September 1993 - May 1994.

- i) Statistically analysing the collected data using Factor Analysis.
- ii) Reporting the findings.

3.2 Research Procedure 1: Piloting the Comprehension Tests in Nottingham

Prior to the piloting procedures, several important steps were taken into account in the attempt to replicate the study in Malaysia. All the texts were taken from Lunzer and Gardner (1979) and translated into L1 by the researcher and later subjected to rigorous checking in terms of the appropriateness of the structure of the targeted language and meaning. The four translated comprehension texts and questions were properly checked by a senior linguist in Bahasa Malaysia from the Department of Education, International Islamic University of Malaysia. A few modifications were purposely made to the four original English texts with the view that the sociocultural norm and background knowledge of the targeted students should be taken into consideration. In a sense, an appropriate and balanced perspective was the guiding principle in replicating Lunzer and Gardner's (1979) study in a Malaysian context. Any information from the original texts which could sound alien and could have caused problems was purposely modified with relevant localised information which was native and matched the possible background knowledge of the students.

As an example, in Greig's original text, the geographical names of the North Sea and Shetland Islands could have disoriented and posed a problem for the targeted students and these names are localised to the more familiar names such as the South China Sea and Titiwangsa Mountain Ranges. As well as changing the geographical names the researcher made several other changes to the original texts and this was done with the view of accomodating the sociocultural differences of the targeted students. As an example, in Jane Gilbert's original text, Enton was changed to Intan and instead of 'the year of 1900' in the original Jane Gilber's text it was changed to 'fifty years ago' in order to match the equivalent time frame in Malaysia. No changes were made to the original texts of Alistair and Brighty except in changing the name of the actors to more suitable local names; Ali and Si Pintar respectively.

3.2.1 Replicating Lunzer and Gardner's Eight Hierarchical Subskills: Piloting the Modified and Translated Texts in Nottingham.

As mentioned above, this study replicates all the four texts and the comprehension exercises used by Lunzer and Gardner (1979). The texts and exercises are adapted and translated into Bahasa (L1). (All the texts in L1 and L2 and the accompanying comprehension questions for each text can be seen in Appendix C.) In preparing the texts, the researcher gave a lot of consideration to the students' sociocultural background. In all the adapted texts (L2) which are then translated to (L1), the researcher devoted great attention to the flow, the content and appropriateness of the targeted language.

All the L1 and the L2 texts were piloted on 3 Malaysian teenagers aged between 13 and 15 who were at that time accompanying their parents studying at the University of Nottingham. The piloted students were studying at local secondary schools in Nottinghamshire. They were tested on the comprehension exercises and later interviewed on several aspects of the texts and the tests' questions such as the suitability of the words, phrases and sentences expressed particularly in the L2 texts. The interviews were open-ended and some changes were later made to the texts.

As well as the three students, it is overwhelmingly important to discuss the responses gathered from four teachers on the appropriateness of the texts. Four language teachers were consulted on various aspects of the translation and the suitability of the texts and the comprehension exercises. At the time of the piloting, the teachers were studying for their Teaching English as a Second Language (TESL) Bachelor Degree at the University of Nottingham. They had vast experience in teaching language in primary and secondary schools in Malaysia. These teachers were given enough time to verbally reflect on the suitability of all the texts and the comprehension exercises. Later, the verbal inputs gathered from the teachers were further used in making a few changes to the texts and the test items. In principle all the teachers agreed on the modifications and the translations of the texts particularly on the suitability of the texts and the test items for the targeted students. They also agreed that the

translated texts did not deviate from the original texts particularly in terms of the intended meaning.

The scores obtained by the teachers after the first piloting of the comprehension tests can be seen from the following table.

Table 2: Pilot Results of the Comprehension Tests

Teacher	Text	Language	Score
A	Ahmed Idris (Greig)	Bahasa	31/36
	Ali (Alistair)	English	36/41
B	Si Pintar (Brighty)	English	36/41
	Ali (Alistair)	Bahasa	36/41
C	Ahmed Idris (Greig)	English	28/36
	Maimunah (Jane Gilbert)	Bahasa	30/36
D	Maimunah (Jane Gilbert)	English	35/36
	Si Pintar (Brighty)	Bahasa	40/41

Each teacher was provided with all eight texts and the comprehension questions. For every pair of texts the teachers' tasks were to read the passages, to find answers to the comprehension exercises and to comment on any deviation of meaning observed. As we can see from Table 2, the variations of the scores of correct answers across the four teachers highlighted some possible problems with the texts and the comprehension questions. The scores were not intended to be used as a guide in rating the comprehension exercises along the difficulty continuum. They were used as a general guide in identifying questions that had caused problems to the piloted teachers. A question may well be too easy for one reader but a problem to another reader. With this in mind each teacher was open-endedly interviewed with the hope of highlighting the

possible obstacles and the teachers' comments were used to modify the texts and the tests. The researcher observed that most of the comments made by the teachers had to do with the choice of words and phrases in L2. This problem was well anticipated.

In essence, the pooled students and teachers' judgements were used to guide the researcher in providing a balanced replication of the modified texts and indirectly predicting the readability of the texts and the comprehension questions.

Jane's text was an original work of Lunzer and his team. *Brighty* and *Alistair* were adapted from the Scott Foresman Reading Systems Programme. As for *Greig*, it was adapted by Lunzer and Gardner (1979) from Gough, C. (1960) *Boyhood of Great Composers* London: Oxford University Press. At the beginning it's hoped that *Greig's* (Ahmed Idris) factual text (tested on 199 children) from Lunzer and Gardner's study would yield reliable results especially in identifying and proving the hierarchy of subskills. It failed. They thought that the small number of items from each category of the eight subskills of the test plus the simplicity of the test itself could be the reasons for the scarcity of evidence for the existence of the subskills. Following those possibilities, a literary text, *Brighty* (Si Pintar), was adapted from the Scott Foresman *Reading Systems Programme*. Nearly 300 children took the test and analysing the data yielded only 3 factors. Out of these 3 factors, one factor was thought to load on the items tested from both tests. Fearing that the texts' types might '...mask the differentiation into item types for which evidence was sought.' (Lunzer and Gardner, 1979, p. 45) another two identical texts were made to match the first two texts. So, *Alistair* (Ali) was adapted from the Scott Foresman Programme. *Jane's* text (Maimunah) resembled the *Greig's* (Ahmed Idris) text. All the texts have equal content, interest level and format.

According to Lunzer and Gardner (1979) many revisions (in terms of the facility level, ease of marking, level of distractors, etc) and trials were made on a sample of 12-15 children. The children and the teachers were involved in the

discussion of the items. The number of versions for *Greig* was at least 10, *Brighty* 8, *Alistair* 5 and *Jane* 3. The complexity of the original study is worth mentioning and comparatively although this replication study may not have equally experienced the lengthy processes of producing the appropriate version as in the original study it has experienced the crucial processes of gaining the reliability of the L1 and the L2 tests after the lengthy, fruitful experience of translating and adapting the texts for the targeted students.

3.3 Research Procedure 2: Testing the Comprehension Subskills in Malaysia.

Introduction

Prior to the testing, some 300 primary and 150 secondary students were identified as the potential sources of data. The relevant authorities in each school had given their assurances that the students were able to read and write accordingly. Evaluating the samples yielded enough valid responses to carry out an effective Factor Analysis; 271 students from the primary schools were tested on the L1 comprehension tests and 138 students from the secondary schools were tested on the L2 comprehension tests.

3.3.1 Testing the L1 and L2 Texts on the Primary and the Lower Secondary School Pupils

A selected number of students (n=409) were identified from four urban schools in the district of Johor Bahru, the metropolitan-capital state of Johor, Malaysia. According to one of the district educational officers, these inner-city schools are categorized as average schools in terms of the overall academic performance. Some of the schools have a population of more than 1500 students.

The randomly selected samples were identified from the four schools. 300 students from two primary schools and 150 students from two secondary schools responded to the four texts. The students, aged 12 and 15, were required to read and write the answers to the questions on the four separate comprehension texts which were replicated from Lunzer and Gardner's (1979) original works. At the outset it is important to note the different entry ages for primary schooling in Malaysia and England; age 5 in the U.K. and age 7 in Malaysia. The texts and the comprehension exercises (See Appendices C(i) and C(ii)) for the primary students were translated into L1 with a few modifications made to the original texts derived from Lunzer and Gardner's (1979) study. This was done with the view of suiting the pupils' socio-cultural backgrounds. Despite the modifications,

the gists, contents and flow of the stories are adequately maintained so as to avoid or at least minimise unnecessary biases when comparing the results with those of Lunzer and Gardner (1979). The readability of the texts was also taken into consideration.

Later, the responses were coded and subjected to a series of data analysing techniques as in Lunzer and Gardner's (1979) work. Factor analysing the data yielded 409 reliable responses, which is well above the required standard minimum sample size of 200. Rost (1989) says '...only samples of N 200Ss can yield satisfactory, sample-invariant solutions...large enough to reduce the standard error of the correlations to negligible proportions.' (p. 91).

In general, each text has about thirty questions which are categorised into eight groups of subskills. Lunzer and Gardner (1979) identified the eight comprehension subskills as listed in Chapter 1.3. The simple explanation of the given meanings for the eight subskills are taken from Lunzer and Gardner (1979) and Waite (1980). It is also important to note that the number of questions asked for each category of skills is adequately balanced throughout the comprehension tests. The following Table 3 (see page 76) outlines a detailed distribution of each tested subskill and the corresponding questions throughout all eight texts.

Table 3: Distribution of types of question according to the eight subskills.

	TEXTS			
	Ahmed Idris (Greig)	Maimunah (Jane Gilbert)	Ali (Alistair)	Si Pintar (Brighty)
Hypothesised hierarchy of subskills and the corresponding questions				
Word Meaning (W)	1a,1b,1c, 1d (4mks)	1a, 1b, 1c 1d (4 mks)	1a, 1b, 1c. (3 mks)	1a, 1b, 1c (3 mks)
Words in Context (WIC)	2a,2b,2c (3 mks)	2a,2b,2c (3 mks)	2a,2b,2c (3 mks.)	2a,2b,2c (3 mks)
Literal Comprehension (L)	3,4,5,6,9 (5 mks)	3,4,5,6,7 (5 mks)	3,4,5,6,7 (5 mks)	3,4,5,6,7 (5 mks)
Drawing Inferences from Single Strings (ISS)	7,8 (2 mks)	8,9 (2 mks)	8,9,10,11 (4 mks)	8,9,10,11 (4 mks)
Drawing Inferences from Multiple Strings (IMS)	10,12,13i, 13ii (4 mks)	10, 11, 12i, 12ii (4 mks)	12,13a,b,14 15&16 (8 mks)	12,13a,b 14,15,16 (8 mks)
Interpretation of Metaphor (M)	15a, 15b (2 mks)	13a,13b (2 mks)	17a,17b,18,19 (4 mks)	17a,17b 18,19 (4 mks)
Finding Salients or Main Ideas (S)	17,19 (7 mks)	14,15 (7 mks)	20,21 (9 mks)	20,21 (9 mks)
Forming Judgements (J)	11,14,16,18, 20 (9 mks)	16,17,18,19 20 (9 mks)	22,23,24 (5 mks)	22,23,24 (5 mks)
Total Number of Questions	27	26	29	30
Total Possible Marks	36	36	41	41

Prior to the comprehension tests several important consultations were held with the affected local teachers on the readability of the texts. As for the secondary students, the texts were written in the English language with some adaptations to the socio-cultural backgrounds of the readers. In Malaysia, English, as a second language, is taught in both primary and secondary schools. Prior to the testing, some 300 primary children and 150 secondary students were identified as the potential samples and the relevant authorities in each school had given their assurances that the students were able to read and write in English accordingly. Evaluating the samples yielded enough valid responses for Factor Analysis; 271 from the primary schools and 138 from the secondary schools.

A plan for collecting the data had been properly designed and this is best illustrated in Table 4.

Table 4. Samples under study: 450 students (aged 12 & 15) from 12 classes in the four inner-city primary & secondary schools.

	Test Session			
	1	2	3	4
Primary School A	B1*	B2	B3	B4
Primary School B	B4	B3	B2	B1
Secondary School A	E1*	E2	E3	E4
Secondary School B	E1	E2	E3	E4
Note: B* connotes Bahasa (L1) texts (native language texts) E* connotes English (L2) texts (second language texts)				

The English texts replicate Lunzer and Gardner's (1979) work in England but were tested on a different age group in Malaysia. Lunzer and Gardner (1979) tested the texts on children aged 10-11 years. The initial intention was to test the English texts on the same age group in Malaysia. This plan had to be abandoned and some alteration was unavoidable, mainly as a result of prior discussions with the affected teachers as to the readability and suitability of the texts. They spoke out their experiences and believed that the L2 texts matched the 15 years old

better. Another language teacher expressed her view that the L2 texts may be on par with less than 20 good pupils in her school. Her view was taken into consideration but at least she acknowledged the reliability of the texts for the targeted pupils.

Although much assurance was given to the teachers as to the readability and the reliability of the tests, they expressed the fear that the length of the L2 texts would mean more reading time than the L1 texts and subsequently this would consume more class periods and could easily disrupt the already packed teaching schedules. To conduct all of the tests would consume five to six class periods. The readability of the texts was very much anticipated prior to the testing and was taken care of during the piloting process. Although the tests had been proved readable through the piloting process in Nottingham a compromise had to be reached. The prior planning had to be adjusted and a new plan was laid out as seen in Table 4. The teachers agreed on the new plan and subsequently the potential primary school samples were identified to sit for the L1 tests. So, the whole issue of not allowing the L2 texts to be tested had to do with the reading time factor and the very cramped language syllabus that had to be delivered before the primary school standardised assessment test.

In a sense the work of Lunzer and Gardner (1979) was very much replicated in Malaysia. The readers in both countries sat the tests written in their first language. What is unique about this study is that it extends the scope of the data source to a second language. English is considered as a first language in schools in England but as a second language in Malaysian schools. The lower secondary schools samples, aged 15, were tested on the L2 texts. Through the consultations, the head teachers and the language teachers of the secondary schools strongly recommended the L2 texts for the aged 15 groups.

Throughout this study the following codes are used to signify the corresponding texts for all the tests in L1 and L2. The letters 'B' and 'E' are used to represent the L1 and L2 texts respectively.

<u>Code</u>	<u>Title</u>	<u>Topic</u>
B1 & E1	Ahmed Idris (Greig)	biography
B2 & E2	Si Pintar (Brighty)	narrative/adventure
B3 & E3	Maimunah (Jane Gilbert)	biography
B4 & E4	Ali (Alistair)	narrative/adventure

The allocated time given for each test was about 45 minutes. Occasionally, a few of the students were permitted to work slightly beyond the limited time frame.

It is important to note that the distance between the two primary schools is about 6 kilometers. Table 4 was constructed with the aim of avoiding and possibly minimising potential discussions of the texts among the students since they were tested in different time frame. Furthermore, the catchment area for each of the schools is not the same. Primary School A is 3 kilometers to the east of the town and Primary School B is 3 kilometers north of the town. Their situations reduced the possibility of discussion among the selected students. The secondary schools are adjacent to each other. However, the execution of the four tests in each school did not take place concurrently. The samples from secondary school A were tested in the morning and in school B the tests were done in the afternoon.

All the schools can be categorised as “average” in performance, but despite the generalised categorisation there are readers who are excellent in their academic achievements. This reduces the selection of samples as being biased and it colours the samples with a mixture of different achievements or abilities. The four standard six classes in each of the primary schools were mainstreamed accordingly; class number 1 being the best. The mainstreaming of the children is believed by the teachers to be the best way to tackle some of the learning

problems faced by the children. Furthermore, the children are expected to sit for the standardized examination known as Primary School Evaluation Testing at the end of their primary schooling. The test is one of the facets of evaluation introduced by the Kementerian Pendidikan Malaysia (Ministry of Education) in assessing students' performances. It is based on the Integrated Curriculum for the Primary School. This new innovation was introduced by the authorities about a decade ago in an attempt to produce better students; physically, spiritually and intellectually. In primary school A, about 90% of the total standard six students were identified as the samples and for primary school B, almost 100% participated in the tests. It was anticipated that the execution of this study was at the prime time of the children's mastery of whatever language skills they had acquired.

The secondary schools are not mainstreamed and the population for each class is heterogeneous; a mixture of different language performances, racial, economic and social backgrounds. The lower secondary school samples were to sit for the Lower Secondary School Assessment tests a month after the conduct of the data collection. In a similar fashion as the primary schools pupils, the lower secondary schools students were also at their prime time of learning and mastering the L2 skills.

The following chart outlines the general reading components of the primary school system which are in operation in Malaysia. It is not possible to outline the details of all the skills taught. The chart which is specified for the English language reflects similar reading skills emphasised in the Bahasa syllabus.

Chart 1: English Language Reading Components for Malaysian Primary Schools

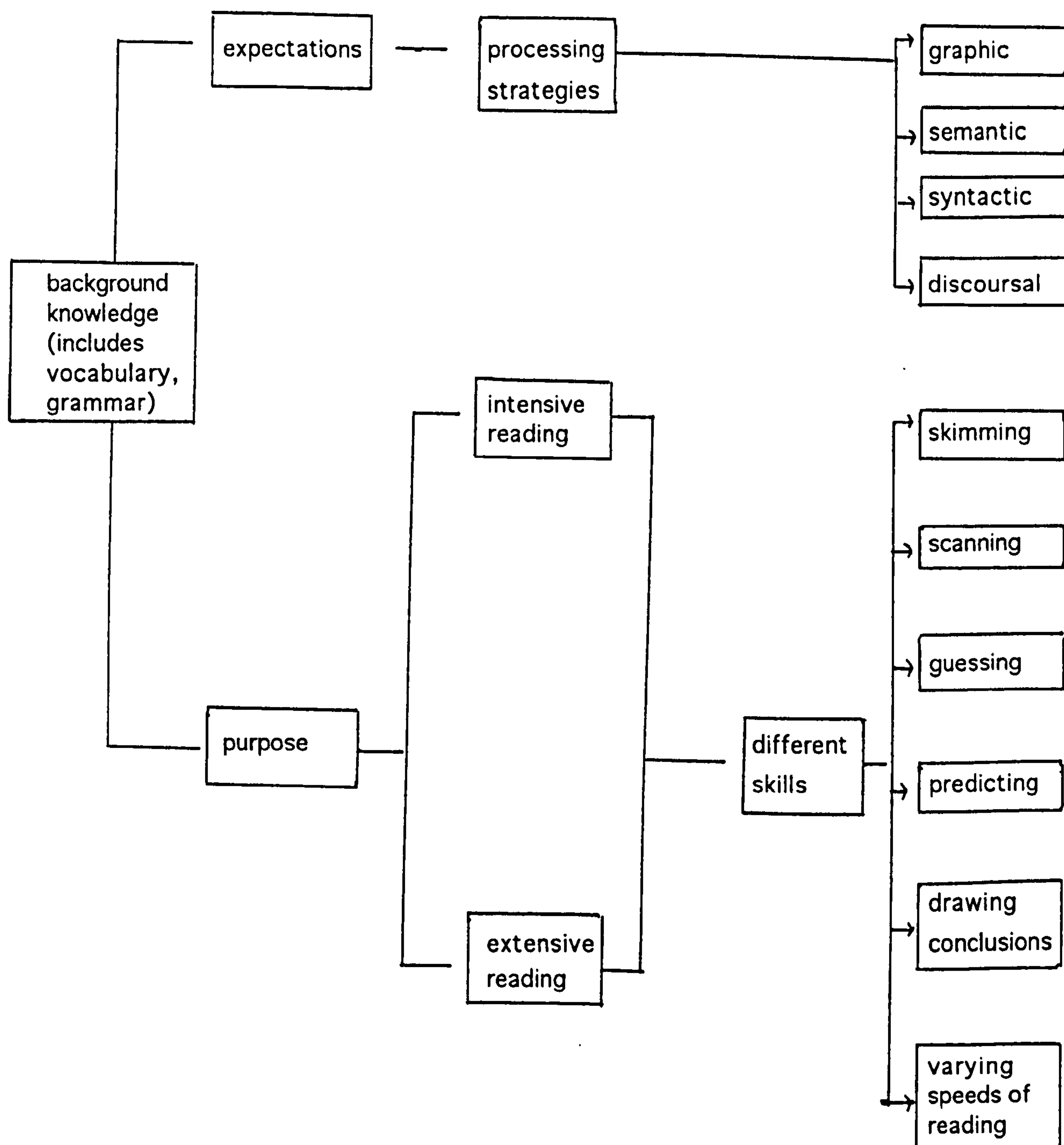
<u>Phase</u>	<u>Age</u>	<u>Reading Skills Taught</u>
I	7-9	emphasis on aural-oral skill and later on reading and writing. Reading skills: word recognition, phonics and comprehension.
II	10-12	consolidating the four skills in communicative behaviour. Developing reading skills, reading habits, study skills; intensive & extensive reading, gathering information from various sources, identifying cause and effect, sequences of ideas and predicting results.

(Based on the ‘Special Guide Book for English Language Year 5’; Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia, Kuala Lumpur 1990.)

What can be judged is the fact that Phase II of the English reading skills taught in the primary schools would be appropriate for the L2 texts and this fact had been the guiding principle in choosing the appropriate age group for the tests. As mentioned earlier this plan had to be abandoned in the process of reaching a compromise. Still, whatever reading skills are learned in L2 may be to some degree transferred to the L1 tests. The transfer of such skills is not the focus of this study.

As for the Bahasa reading programmes, the general reading components objectives for phase 1 are to produce students who can read with correct pronunciation and intonation and are able to read and to understand basic reading materials. The fact is that the reading components for phase 2 are more challenging than in phase 1. Students are taught to understand, summarize, appreciate, enjoy and to interpret information from a variety of genres. Thus, the testing of the L1 texts to the primary schools samples was adequately justified by the stated reading syllabus. (Ref: **Special Guide Book for the Malay Language.(1990) Year 6** (New Curriculum for Secondary School. Dewan Bahasa dan Pustaka. Kementerian Pendidikan Malaysia).

In general, the methodological approach that underpinned the reading syllabuses for both the L1 and L2 reading skills agenda has been communicative. One of the aims is to achieve individual development in a holistic manner; stressing the basic skills of reading, writing and arithmetic. The English Language syllabus for the secondary school acknowledges the four essential practical communicative skills; listening, speaking, reading and writing. It is important to note the integrity of the three main areas of language; phonology (pronunciation), lexis (vocabulary), and grammar. The syllabus stresses the importance of viewing those areas as '*interlocking circles*' and not as '*compartments*'. Clearly, it is the interplay of the four skills in those three areas that manifests the integrated-communicative syllabus. This syllabus integrates the academic/'ivory tower' and the practical/communicative/'market place' approaches. Thus, the best features of the intellectual and the practical aspects of language are amalgamated to form the concise eclectic approach. At this junction, an overview of the secondary school's English language reading syllabus is best illustrated by the following flow diagram:



(The secondary school's discussion and figure 4.1 are taken from 'Compendium: A Concise Guide For Teachers of English.' Published by the English Language Programme Committee, Malaysian Ministry of Education; 1989a, p. 39.)

From the above figure 4.1 it is clear that the secondary schools' students had been exposed to such skills and this brings the idea that the tests are valid and reliable as far as the students' learning of the subskills. A further related phenomenon of the

reading skills taught in the secondary schools can be seen in Appendix A which gives a detailed outline of the reading objectives for form III in Malaysia.

DATA ANALYSIS

3.4 The Data Analysing Techniques

The purpose of this study was to identify the existence of distinct subskills in silent reading comprehension tests by replicating and extending the work of Lunzer and Gardner (1979). A total of 500 children, aged 12 and 15, were empirically tested on the eight hypothesised subskills. The correlations of their scores or results were analysed by the SPSS-X RELEASE 3.0. package. Principal-Components Analysis (PCA) and Factor Analysis (FA) are specifically chosen in analysing the data. Youngman (1979) suggested that the above techniques are effective in identifying and understanding patterns of relationships when a large number of correlated variables are involved. These two analytical techniques are chosen for their abilities to extract the hidden values, also known as components or factors, from a large number of variables. In general, the analysing procedures are as follows:

- a) finding the correlation matrix of all the eight variables scores.
- b) 1-tailed significant of correlation matrix--Pearson-Correlation Coefficients.
- c) Principal-Components Analysis (PCA).
 - i) initial statistics--eigenvalues.
 - ii) factor matrix--extracting the factors.
- d) Factor Analysis. Kaiser Normalization--Oblimin Rotation.
 - i) rotated factor matrix.
 - ii) factor transformation matrix.

3.4.1 Factor Analysis and its Background. From Thurstone (1947) to Lunzer and Gardner's (1979) Work.

The following discussion is an attempt to elaborate more on the choice of Factor Analysis (FA) as the statistical means and to explain the statistical data presented in part I of this study. In doing so, the basic principles of the FA are

discussed and integrated with the presented data and it is hoped that the attempt to explain and understand the important processes involved in analysing the data will bring a clearer picture as to the complexities of this study .

The nature of this study could be regarded as nonexperimental. To be more specific there was no control group involved in the testing and finding of the relationships among the eight variables taken from the 500 samples. Factor analysis (FA) and Principal Components Analysis (PCA) are the chosen techniques for identifying the independent variables out of the set of the eight variables. In general, PCA diagnosed the nature of how the variables scores clustered together and later produced a small number or set of characteristics of uncorrelated components which reflect the correlations among the tested variables. PCA is executed prior to the FA. As for the FA, Tabachnick and Fidell (1989, p. 28) suggest 'When there are hypotheses about underlying structure, FA is often used to develop the structure and assess the fit between the data and the hypothetical structure. In this case the underlying IVS (*independent variables*) are called factors.'

In this study, the attempt to find the underlying factors that could contribute to the formation of the hierarchy of subskills is a complex process. PCA and FA are the best methods of bringing to the surface the underlying factors or values. In a sense, as stated by Tabachnick and Fidell (1989, p. 597) PCA and FA are:

...applied to a single set of variables where the researcher is interested in discovering which variables in the set form coherent subsets that are correlated with one another but largely independent of other subsets of variables are combined into factors. Factors are thought to reflect underlying processes that have created the correlations among variables.

Clearly, PCA and FA are designed to form patterns of correlation from the tested variables and later combine them into some identifiable factors. Thus, the factor analysis used is confirmatory in nature because it tries to test whether the given hypothesis should be accepted or rejected (Kline, 1994). In this study, patterns or factors within the eight variables need to be identified. What is important is that the identified values or factors can be related to the reading comprehension theories and hypotheses as discussed before.

FA is known as a powerful statistical method. Historically, according to Child (1970), Francis Galton was the pioneer of the factorial study. About a hundred and twenty years ago this scientist laid the foundations of FA with two major ideas which are best explained by Child (1970, p.3) as consisting of:

...general intellectual power...spread in a continuous fashion from the very dull to the very bright...[and] the concept of correlation. He developed quantitative methods to give some idea of the interdependence between two variables;...

Weiss (1995, p. 792) noted that as well as the concept of correlation, Galton also discovered the concept of '...regression during experiments with sweet-pea seeds to determine the law of inheritance in size.' Regression, also known as multiple correlation, is a measure of the '...degree of association between three or more variables simultaneously.' (Cohen and Manion, 1989, p. 158).

Subsequently the two ideas developed and early in this century Karl Pearson, a famous British statistician, championed and extended the correlation concept and to what is known today as the Pearson product-moment correlation. In general, Kline (1994, p. 18) says 'correlation is a measure of the degree of agreement between two sets of scores from the same individuals.' Kline (1994) said that in the early of this century Spearman produced a report on human general intelligence and the report was based on the concept of exploratory factor analysis in the quest of finding the main dimensions or constructs of the

positive correlations of human abilities.

In 1947 L. L. Thurstone published a work entitled *Multiple-Factor Analysis*. The work was done in the Psychometric Laboratory at the University of Chicago. It explained the complexities of the iterative computation of communalities (h^2) and the rotation of factors in finding the simplest explanation (law of parsimony) of the correlation out of the many rotations. The iterative computation of communalities is not without problems and Kline (1994) noted three main associated problems that researchers should be aware of in producing the number of factors out of the factor analyses and the factor loadings produced from the iterative methods. Factor loadings should be interpreted like correlation coefficients and in a way a loading in a factor is a representation of a correlation between a test and the factor. Thus Kline (1994, pp. 45-56) observed that the computational results should not be accepted as absolute truths because of the following reasons:

1. It is assumed that the number of factors emerging from the first iteration is correct, for this number is extracted in all further iterations. There is no proof that this is indeed the case.
2. Furthermore the first estimate that was put in diagonals of the correlation matrix actually affects the final solution, again casting doubt on the method. Indeed it must be realized that there is an inextricable link between the number of factors extracted and the communalities (Cattell, 1978).
3. Finally iterative methods can lead to communalities greater than unity which makes no sense.

In its simplest sense, a communality or h^2 (h^2) is actually a representation of the sum of squares of the factor loadings for every single variable tested. In general, a communality is interpreted as having a value between 0 and 1. A value of communality 0 for a variable reflects that the factors shared

no sameness or nothing in common with the variable and as the figure comes close to 1 the degree of sharedness between the variable and the factor becomes greater (Wright and Fowler, 1986).

3.4.2 Performing FA and Its Related Problems and Issues

In this study the reading comprehension data are very complex and the raw data must be reduced to some meaningful factors if one wants to understand the relationships among the factors. FA is chosen as the method of analysis because of it is robust and among others, according to Lunzer and Gardner (1979, p. 51) it helps to:

- a) summarise a set of correlations between tests through the clustering behaviours and
- b) summarise individual performances '...on every one of the measures...by calculating [individual] performance on each of the underlying factors.'

The objective of performing FA is to find as many factors as possible from the tests that measured the eight variables. The aim of FA is to find the shared variance because from this variance the underlying factor or factors could be extracted from the many tests. But as well as this, FA needs to analyse the common variance, specific variance and the error variance of all the tests (Hatch and Lazaraton, 1991).

Lunzer and Gardner's initial hypothesis was rejected by using the exploratory factor analysis method. Lunzer and Gardner (1979) found no subskills as such. In this study, the methods of analysis used by Lunzer and Gardner (1979) are replicated. In doing so, several steps are cautiously taken before reaching the final conclusions. The following discussion is pertinent before FA can be executed.

Finding the reliability of the tests is done by finding the consistency of the tests. In this case, the four tests in L1 and the four in L2 are subjected to item analysis. In its simplest sense, Kline (1994, p. 127) defines item analysis as follows:

Items which correlate the most highly with the total score are selected on the grounds that these are measuring what most of the items are measuring. Thus the set of item so selected ought to be virtually identical to the set of items loading on the general factor. In addition a further criterion for item selection is that the correct answer should be obtained by 20-80 per cent of the sample.

Lunzer and Gardner (1979) also acknowledged that the majority of the F values in the item analysis should fall between the 20 to 80 range. Table 5 summarised the behaviours of F values (the percentage of correct answers) taken from the study in Malaysia which is then compared with Lunzer and Gardner's study (1979) in England.

Table 5: A comparison of the number of items from each language test falling into the percentage of the right answers bands

	Present Study in Malaysia. Bahasa Tests English Tests Table 6 Table 7		Previous Study in UK. English Tests. Taken from Table 3.2 in Lunzer & Gardner (1979, p. 50)
% of Right Answers' Band			
20-80	74/114	90/114	97/112
35-65	39/114	47/114	59/112

The comparison of the Bahasa and the English tests from the above table provides an important finding; that the Bahasa tests are comparatively less reliable than the English tests tested in either Malaysia or in England. No doubt this result would have a considerable impact on the factor analysis of the Bahasa tests. Although the tests are very similar in the sense that they are a direct replication of Lunzer and Gardner's study (1979) in England, there are other delicate matters that probably could have caused such differences.

Part of the answer to the problem has to do with the fact that the attempt to translate one language into another and keep the translated language at the same level of translation is not without underlying problems. When we translate a text to another language the difficulty level will change too, in that some items will be difficult and others will be easier. So, translation changes the text and therefore the individual difficulty level is changed. The comparison of the number of items falling into the percentage of the right answers' band as presented in Table 5 is interesting in the sense that it is not a weakness but rather a difference. The problem is worth reporting because the tests are reliable but the results of the L1 tests in Malaysia are not the same or close to the English tests' results observed from Lunzer and Gardner's study in England. The passages and questions for all the tests are the same. As mentioned before, some adaptations of Lunzer and Gardners' original texts are inevitable. This is to make certain that some aspects of the source language (in this case Lunzer and Gardner's texts) are properly adapted and translated into the targeted language so that the important meanings and messages are not lost and can be clearly understood by the readers. This problem is discussed by Steiner (1975, p. 32) as involving some elements of translation such as '...the uses of inflection, grammatical structure, and word-choice by different social classes and ethnic groups...' that need to be fairly addressed. Throughout the process of adapting and translating the original texts the researcher had to rationalise the interpretation of the texts.

This hermeneutic or 'understanding of understanding' concept of translation (Steiner, 1975, p. 414) is not without related issues. Newmark (1989, p. 13) puts forward questions regarding translation such as:

Is the meaning to transfer the meaning intended by the writer, or to reformulate the meaning intended by the translator? Is it to be modified for the reader, or again is it to be squared with the facts of the matter? There is no straight answer - it depends on the purpose of the translation.

Such complexities are encountered at the beginning of the translation process. What is crucial is that the appropriateness of the translation was judged to be proper and adequate by an expert in Bahasa from Malaysia. At the outset of the analysis of the data from the Bahasa texts is the fact that the texts should have been re-tested a couple of times until they matched the needed results. But equally important is the fact that if this testing and re-testing had been done, it would have been different tests. It would be more difficult to revise the Bahasa texts a couple of times and very likely that further revisions would cause further differences from the original texts. What is at stake is the fact that the tests were trying to replicate the UK's study, so the Bahasa and the original versions should be broadly parallel in content. This issue of comparability means that the more changes made to the Bahasa texts, the more difficult it would be to make the claim that the Bahasa texts are parallel with the original versions. In other words, the items' content would have to change and thus affect the items' difficulty.

Two texts are culturally and ideologically transformed in order to confront any possible dispute in terms of the reliability of the tested texts. The autobiographical texts of '*Greig*' and '*Jane*' are transformed into '*Ahmed Idris*' and '*Maimunah*' respectively. No cultural or ideological changes are made on the original narrative texts of '*Brighty*' (*Si Pintar*) and '*Alistair*' (*Ali*). In other words, the narrative contents of the original texts are well maintained.

Further refining of the Bahasa texts in trying to match the reliability figures of the English texts is not practical and would distance the texts further from the original intention of directly comparing the translated tests with Lunzer and Gardner's English texts tested in the UK.

The above issues surrounding the Bahasa texts are also accompanied by other constraining factors in trying to test the texts in the selected schools. In most cases, it is not easy to slot into the fully packed schedules of the schools' timetables. At the time of testing, the teachers were busy preparing the students for the national assessment held in the beginning of October every year. Testing and re-testing the texts in the same schools would reduce the reliability because of the uncontrolled factors such as discussion of the tests among the tested students after the testing.

Why test for reliability? It is pertinent that the results of the tests can be judged as consistent or stable. In other words the tests should measure consistently whatever it purports to measure. In this case, the study is governed by two contrasting hypotheses: reading comprehension consists of a hierarchy of subskills which can be distinctly identified. Or, alternatively, reading comprehension constitutes a unitary skill. All the comprehension questions are designed in a hierarchical fashion of difficulty. The following paragraphs discuss a few methods of finding the reliability of the tests.

3.4.3 Finding the Correlation Matrix by Item Analysis.

By definition, a correlation matrix '...is a set of correlation coefficients between a number of variables.' (Kline, 1994, p. 4). In this study hundreds of correlations are produced from the eight variables. To simplify the figures, item analysis is used to clarify these huge correlation figures into interpretable information. The purpose of item analysis is to find the correlation of each test item with the total score of the test. Kline (1994, p. 127) explains 'Items which correlate the most highly with the total score are selected on the grounds that these are measuring what most of the items are measuring.'

Lunzer and Gardner (1979, p. 49) say 'Items with very low or very high Fs (% of correct answers) rarely achieve a high r [correlation]. ' Parallel tests were conducted on the students to ensure the consistency of the tests. In other words, each student was subjected to four different tests. Now, reading the index (or

coefficient) of correlation entails the understanding of the value range of +1.00, perfect positive correlation and -1.00, perfect negative correlation with 0 [zero] meaning no correlation at all. The idea of correlation is to see whether two or more variables correspond. In this research, the two variables are the items that the students got right and the scores on the total test. Before going any further it is wise to understand that there are five data requirements that should be fulfilled before the Pearson product-moment correlation can be executed. Clegg (1990, p. 183) in his operation schedule of the correlation mentions the following requirements:

1. Scores for comparison must be paired off in some manner. Usually they will have been obtained from the same source, but this may not always be the case.
2. The relationship between the two variables must be a linear one. This can be assessed from a scattergram.
3. The scores must be of at least interval status.
4. The scores must be normally distributed.
5. The two sets of scores must have similar variances.

Table 6: Item analysis for four Bahasa tests (n=271) Age 12

	Ahmed Idris (Grieg) Test 1		Si Pintar (Brighty) Test 2		Maimunah (Jane) Test 3		Ali (Alistair) Test 4	
<u>Subskills</u>	F *	r **	F	r	F	r	F	r
(W)	86	36	78	18	76	29	89	21
Word	52	33	95	30	82	27	38	22
Meaning	87	40	93	38	46	30	40	19
	21	15			53	23		
(WIC)	54	38	69	35	85	36	61	16
Words in	86	38	41	31	69	33	30	20
Context	46	32	95	37	74	38	91	26
	93	32	90	33	76	28	79	34
(L)	60	39	45	26	96	35	78	36
Literal	68	39	90	37	97	44	88	36
Comp.	41	28	92	27	87	29	93	29
	93	48	87	37	92	41	93	36
(ISS)	85	39	22	18	93	25	10	16
Inferences	27	27	22	31	86	38	07	31
Single			61	40			45	24
Strings			87	24			45	33
(IMS)	93	40	17	06	52	45	76	22
Inferences	35	-00	27	34	37	20	34	15
Multiple	96	32	26	36	85	46	18	14
Strings	90	37	40	27	62	41	43	18
			56	18			76	44
			34	33			38	17
(M)	04	30	26	37	76	36	17	29
Interpretation	02	13	39	46	10	14	57	39
of			24	29			30	44
Metaphor			55	24			42	28
(S)	63	54	74	48	63	39	49	45
Salient or	38	24	50	46	48	25	76	57
Main Idea								
	78	38	65	39	79	35	78	42
(J)	09	15	40	21	77	27	89	38
Forming	55	48	40	35	45	18	79	48
Judgement	43	29			32	05		
	78	35			75	43		
* % of correct responses ** correlations of item success with score on total test (e.g. value of .77 as correlated to the perfect value of 1, thus .77 as highly correlated), decimals omitted								

Table 7: Item Analysis For Four English Tests (n=138) Age 1 5

	Ahmed Idris (Grieg) Test 1		Si Pintar (Brighty) Test 2		Maimunah (Jane) Test 3		Ali (Alistair) Test 4	
	F *	r * *	F	r	F	r	F	r
<u>Subskills</u>								
W	73	66	75	46	91	32	61	70
	82	42	57	63	53	49	75	59
	79	40	70	62	77	54	73	32
	20	11			63	51		
WIC	54	22	54	59	54	38	67	65
	72	32	11	-29	84	36	27	29
	25	05	79	43	52	19	12	-06
L	97	-00	80	36	75	29	88	32
	91	41	50	14	99	02	77	60
	79	40	72	60	98	21	83	40
	40	43	67	54	90	51	80	54
	94	23	46	68	97	20	85	51
ISS	49	44	17	19	44	47	86	52
	59	38	47	64	88	54	12	42
			46	29			33	48
			62	48			41	36
IMS	75	64	41	21	75	63	47	41
	58	29	43	66	57	22	49	71
	88	55	51	77	83	64	42	65
	78	56	41	02	82	57	49	27
			47	26			49	32
			21	31			31	11
M	01	10	23	23	49	57	38	62
	36	57	61	67	59	57	57	80
			28	39			11	18
			70	29			51	31
S	74	48	74	74	76	45	61	63
	48	30	55	72	50	30	70	69
J	36	50	56	52	67	70	79	54
	68	69	32	32	64	38	75	36
	73	56	39	35	45	27	73	49
	40	39			42	41		
	79	44			77	64		
* % of the students with correct responses (Facility level)								
* * correlations of item success with score on total test, decimals omitted								

3.4.4 Item Analysis: Interpreting Its Correlation Coefficients

Chapter 4.1.1. elaborates the proper steps in interpreting the correlation coefficients taken from Tables 6 and 7 (see pages 111-119).

Table 8: Discrimination of the number of test items into correlation bands.

	Total number of items in the present study (Malaysia)		Total number of items in Lunzer and Gardner (1979) (England)
Correlation Band	Bahasa Texts (L1)	English Texts (L2)	English Texts (L1)
0.00 - 0.39	93	48	41
0.40 - 0.60	21	42	61
0.61 - 1.00	0	24	10
Total number of <i>r</i>	114	114	112

What can be observed from Table 8 is the fact that the Bahasa texts do not discriminate in the same way as the English texts (L2) tested in Malaysia and in Lunzer and Gardner's texts tested in England. Comparatively the total number of items that are below the .40 correlation is far greater in the Bahasa texts than in the English texts. In this case 93 of the items in the Bahasa texts correlate less than .40 with the overall (total) score and practically this is poor in terms of the two variables. It is also noted that the other two bands of correlation for the Bahasa texts do not match or come close to the corresponding figures from the Lunzer and Gardner (1979) study. This low correlation in the Bahasa texts means that they are not a reliable data source for Factor Analysis (FA). That is why the FA data from the Bahasa texts are not as good as the English (L2) data (see pages 117-118 for other relevant interpretation on Table 8).

The author carried through the analysis of the Bahasa texts but with the knowledge that the expectation of getting the factors from the Bahasa texts was low. If the reliability of the tests is low then the chances of getting stable factors are lower.

But this low reliability is not a problem in the second part of the study. The first part of this complex study makes many deeper and richer new inquiries as to the nature of the reading comprehension subskills issues. Although the FA data from both languages do not provide new insights or conclusions, the study is extended by looking into the students' verbal responses. It does not matter whether there are 'subskills' in terms of processing, what part two of this study shows are the identifiable areas in relation to the discourse of the comprehension. In this, certain readers have certain kinds of discourse patterns and the 'subskills' areas have some utility in discussing those discourse patterns. It does not mean that the first part of the study is concluding that there are 'subskills' as such, but it is appropriate to mention that they are useful applicable labels in discussing the discourse comprehension of the students. The eight 'subskills' are noted to have some stability and robustness and they are useful in that they can be applied in certain ways such as in making certain kinds of predictions about what happens with certain kinds of readers in certain kinds of situations.

3.4.5 Principal Component Analysis (PCA)

As mentioned earlier, it is noted that exploratory factor analysis is the chosen method of analysis used in Lunzer and Gardner's study (1979). It also involved Principal Component Analysis (PCA) and rotating the factor matrix by using Kaiser's varimax rotation in finding the correlated factors. In principle, PCA serves '...to estimate the correlation matrix and this can be done by finding the characteristic equation of the matrix.' (Kline 1994, p. 29). There are two values, called eigenvectors and eigenvalues, which are needed in making the estimation. Now, there are few terms that need more explanation before the meaning of eigenvalues can be clearly understood (see also a slight extended discussion on PCA on pages 125-126).

Variance: A variance is a square of a standard deviation. In a sense, if we want to know students' placements from the mean of an examination, variance can be computed by finding the deviation of each student's score from the mean score. Summing up all the variations of scores from the mean score should total up to zero. Then square the deviation of each score and total up the scores. This total score is known as the *sum of squares*, which is the total variability of the scores' set. The next step is to find the average of the sum of squares (the total variability score) with the number of the students. If the number of students is less than 100 then divide the score with $N-1$. This is to ensure an unbiased estimate of the variance.

Standard deviation (SD): The process of finding a standard deviation is just the same as the variance but it goes a little bit further. Instead of squaring each score from the mean, it is performed by taking the square root of the variance. It is the same as calculating the mean deviation. So, SD is a summary of an average distance of all the scores from the mean of a set of scores. SD is also associated with normal distribution; 1 SD below or above the mean, mode or median score. (1 SD below and above the mean score of 50 with the sd score of 5 means that 68.26% of the total number of scores is between the values 45 and 55.)

So the purpose of performing the PCA is to make some estimations on the correlation matrix. To do this, it requires the identification of the characteristic equation of the correlation matrix which is based on two sets of values known as the characteristic vectors of the matrix (eigenvectors) and the characteristics/latent roots (eigenvalues) (Kline, 1994, p. 29). The correlation matrix (Pearson's product-moment) for the Bahasa (L1) and the English (L2) scores were computed. This initial analysis produced a vast amount of figures and it is not possible to chart the complex intercorrelation among the scores of the eight variables into this explanation. What can be done is to find the highest correlation coefficients in the matrix. By referring to the Appendix B in Child (1970, p. 95), the significance level of the correlation coefficients for the Bahasa scores with the sample size of 271 is set at 0.125 (at 5% level) and 0.163 at 1% level. In this study the coefficients scores are set at 5% level. For the English

scores with a sample size of 138, at 5% level the coefficient is set about the 0.158. So, correlations which are above the indicated levels are worth discussion.

Eigenvalues: An eigenvalue is a statistic which provides an indication of how much variance a factor is taking up. If the factor is made up on the basis of the subskills and if the eigenvalue is less than 1, the factor is judged as contributing less variance than just one of the variables. In other words the factor is so small that it is negligible. What is wanted is an eigenvalue which is greater than one. If it smaller than 1 it is not worth discussing. If the eigenvalue is greater than 1 then the factor might be interesting. This is supported by Kline who says 'The larger the eigenvalue the more variance is explained by the factor.' (1994, p. 30). Refer to the following tables 9 and 10 in terms of the eigenvalues for both the Bahasa and the English tests.

Table 9: Bahasa (L1)

<u>Factor</u>	<u>Eigenvalue</u>	<u>Percentage of Variance</u>	<u>Cum. Percentage</u>
1	6.59	20.6	20.6
2	1.62	5.1	25.7
3	1.45	4.5	30.2
4	1.29	4.0	34.2
5	1.27	4.0	38.2
6	1.21	3.8	42.0

Table 10: English (L2)

<u>Factor</u>	<u>Eigenvalue</u>	<u>Percentage of Variance</u>	<u>Cum. Percentage</u>
1	12.54	39.2	39.2
2	1.52	4.8	43.9
3	1.36	4.3	48.2
4	1.35	4.2	52.4
5	1.19	3.7	56.1
6	1.14	3.6	59.7

In this study, the role of performing the PCA begins with finding the correlation matrix of all the tests. Then, the next step is to find the first principal component that can explain as much as possible of the total variability of the first data. The value given to the first component is called the eigenvalue. In Table 10, the eigenvalue and the percentage of variance of the first factor is almost double when compared with the corresponding factor in Table 9. This is anticipated due to the problems of the reliability of the L1 tests as discussed before. Then, more calculations are made to find the subsequent uncorrelated components that represent the balance of the variances from the data. What can be seen from Table 9 and 10 is a pattern of representation of variances across the extracted components. The new components should represent most of the information of the original data and with these components some possible factors could be extracted in the FA processes.

3.4.6 Kaiser Normalisation (Oblimin Rotated Factor Matrix)

The observed factors from the factor analysis must be rotated if one wants to find interpretable factors. In general, there are two main types of rotations in factor analysis. The first one is orthogonal and it is very difficult to interpret the factors because the factors are uncorrelated or independent of one another. The second one is called rotated or oblique because the factors are correlated and this allows clearer interpretations of the factors (Kline, 1994). In this study, the oblique rotation in the SPSS analysis package is known as *oblimin* method. This method allows three pattern of matrices known as *structure*, *pattern* and *factor correlation* matrices to be produced (Bryman and Cramer, 1990).

3.5 Finding the Reliability of the Eight Tests.

The following information had been discussed in chapter 3.4.2. The author feels that the information discussed before is important in building valid arguments for the reliability of the comprehension tests. Therefore, in the process of justifying the reliability of the data the information must be highlighted again.

There are a number of ways of estimating the reliability of a test. Hatch and Lazaraton (1991, pp. 531-539) discussed four common methods in estimating the reliability of a test. In regard to the four methods Hatch and Lazaraton (1991, p. 531) say you can look for the following:

(1) for consistency over time--correlation between test-retest scores; (2) for equivalence in form--correlation of parallel or comparable tests; (3) for equivalence in judgement--interrater reliability checks; and (4) for consistency *within* a test.

In this study one way of finding the reliability of the comprehension tests is by looking at the percentage of correct answers and the correlation of item success for each of the subskills in the eight tests, four tests in Bahasa (L1) and four tests in English (L2), are presented in Tables 6 and 7 respectively. In Tables 6 and 7 the F index indicates the percentage of right answers chosen by the students and the discriminatory efficiency of every subskill in the eight tests, known as the correlation (r) index, indicates the correlations of item success for the total score of all the subskills.

Lunzer and Gardner (1979) stressed that an effective test is reflected by moderate F-values; neither too high nor too low. Another two indicators of good discrimination in the tests is to observe that '...the majority (of the Fs) should fall between 20 and 80,...and the more items with F values of 35-65, the better the discrimination of the test.' (Lunzer and Gardner 1979, p. 49). The table below is a summary of the percentages of right answers taken from Tables 6 and 7. These

percentages are then compared with the percentage of right answers taken from Table 3.2 (Lunzer and Gardner 1979, p. 50). To simplify the comparison the percentages of right answers are divided into the two recommended bands: 20-80 and 35-65 respectively.

Table 5: A comparison of the number of items from each language test falling into the percentage of the right answers bands

	Present Study in Malaysia		Previous Study in UK
	Bahasa Tests Table 6	English Tests Table 7	English Tests. Taken from Table 3.2 in Lunzer & Gardner (1979, p. 50)
% of Right Answers Band			
20-80	74/114	90/114	97/112
35-65	39/114	47/114	59/112

In Table 5, Lunzer and Gardner's (1979) findings of the 20-80 and 35-65 right answers band are 97/112 and 59/112 respectively (Table 3.2, p. 50) and the figures are compared with the findings from Tables 6 and 7. Clearly, Table 7 (L2 tests) closely resembles Lunzer et al.'s data because the majority of the F values fall between 20-80. Table 6 is close to Lunzer and Gardner's figure by 23 Fs. The discussion of the differences between the Bahasa tests and that of Lunzer and Gardner's tests had been dealt with in Chapter 3.4.2.

Another reliability measure is by looking into the correlation (*r*) of the item with the total score of the test. Lunzer and Gardner (1979) suggest "Those *r* s which fall much below .40 are unsatisfactory...,in the present case at any rate, we would not wish to have too many *r* s which are above .6. This is because, hopefully, the test should also provide differential discrimination for the eight

subareas.” (p. 49). Cohen and Manion (1989, pp. 168-169) suggest four different ranges in interpreting correlation coefficients. They assume that the sample size is more than 100. Correlations ranging from .20 to .35 are regarded as a ‘very slight relationship’ between variables, .35 to .65 as ‘statistically significant beyond the 1 per cent level’, .65 to .85 ‘make possible group predictions that are accurate enough for most purposes’ and .85 and above signify ‘ a close relationship between the two variables’.

If we were to use the Cohen and Manion (1989) categories, the number of *r* s for Tables 6, 7 and Table 3.2 of Lunzer and Gardner (1979, p. 50) could be summarised in the following Table 8a as:

Table 8a: Comparison of the correlation values
(number of *r*'s occurrences) across Tables 6,7 and
Table 3.2 of Lunzer and Gardner (1979, p. 50)

Range of <i>r</i> values Cohen and Manion (1989)	Table 6 (Bahasa L1)	Table 7 (English L2)	Lunzer and Gardner (1979) (English L1)
0-19	17	12	2
20-35	51	27	31
35-65	46	62	73
65-85	0	13	6
85-100	0	0	0
total of <i>r</i> s	114	114	112

It is interesting to note some trends in these r's total occurrences based on the interval r values. Comparatively, the r's occurrences into the five bands for Table 6 do not represent a balanced discrimination between the good, average and poor readers as Table 7 and Table 3.2 of Lunzer and Gardner's (1979) figures represent. Such differences are already discussed in 3.4.2.

Table 11: Reliability of the four tests for the Bahasa Tests
(n = 271). Pearson Correlation's total scores and Cronbach Alpha

1 = AhmedIdris's tests
2 = Si Pintar's tests
3 = Maimunah's tests
4 = Ali's tests

	1	2	3	4	1+2	1+3	1+4	2+3	2+4	3+4	1+2+3+4	Cronbach Alpha
1	-	60	65	58	88	92	88	69	66	70	84	60
2	60	-	62	63	91	68	69	93	91	71	86	63
3	65	62	-	53	71	90	66	87	64	85	82	58
4	58	63	53	-	68	62	90	65	90	90	83	65
1+2	88	91	71	68	-	88	87	91	88	79	95	76
1+3	92	68	90	62	88	-	85	86	72	85	91	75
1+4	88	69	66	90	87	85	-	75	88	90	94	76
2+3	69	93	87	65	91	86	75	-	88	86	94	76
2+4	66	91	64	90	88	72	88	88	-	89	94	78
3+4	70	71	85	90	79	85	90	86	89	-	94	75
1to 4	84	86	82	83	95	91	94	94	94	94	-	86

Table 12: Reliability of the four tests (English Tests, n=138)
(Pearson Correlation and Cronbach Alpha)

1 = AhmedIdris' tests
 2 = Si Pintar's tests
 3 = Maimunah's tests
 4 = Ali's tests

	1	2	3	4	1+2	1+3	1+4	2+3	2+4	3+4	1+2+3+4	Cronbach Alpha
1	-	70	75	75	88	94	91	77	77	80	87	75
2	70	-	75	79	95	77	80	96	95	82	91	84
3	75	75	-	78	81	94	82	90	80	92	89	81
4	75	79	78	-	84	82	96	84	94	96	93	86
1+2	88	95	81	84	-	90	91	95	95	88	97	88
1+3	94	77	94	82	90	-	92	89	84	92	94	87
1+4	91	80	82	96	91	92	-	86	93	95	96	89
2+3	77	96	90	84	95	89	86	-	95	92	97	89
2+4	77	95	80	94	95	84	93	95	-	94	97	91
3+4	80	82	92	96	88	92	95	92	94	-	97	90
1to 4	87	91	89	93	97	94	96	97	97	97	-	94

The reliabilities (the degrees of internal consistency of a test) are measured by Cronbach's Alpha and although high scores of the alpha mean high internal consistencies, the scores that are less than perfect also merit interpretation. Scores of .70 and above are preferable. Furthermore, reliability should be accompanied with validity (Youngman, 1979). The scores other than Cronbach's Alpha are the correlation

coefficients of the total test(s). By mixing different tests in both tests (L1 & L2) the alpha is predominantly higher than .70 and in L2 the majority of the alphas are above .80 which argues a high degree of internal reliability in the tests. Such a degree of reliability is crucial when making a direct comparison with Lunzer and Gardner's study.

In the case of this study a comparable internal consistency of the tests, as is found in Lunzer and Gardner's study (1979), is what is known as Cronbach's Alpha. It is a measure of the internal consistency reliability of a test. Youngman (1979, p. 180) says 'However, it is an estimate and generally only offers a lower bound for the true value. High values are desirable but Cronbach asserts that internal consistency need not be perfect for a test to be interpretable.'

In this study, the Cronbach's Alpha for the Bahasa (L1) and the English tests are compared with that of Lunzer and Gardner's (1979) English tests. By referring to table 11 and table 12, it can be seen that the estimate for the internal reliability of the Bahasa texts is lower than for the English texts. Comparing these figures with table 3.3 of Lunzer and Gardner (1979, p. 52), it is observed that the correlation values of the English (L2) tests are comparable with the English (L1) tests in the UK. The lower internal consistencies of the Bahasa tests than those of Lunzer and Gardner's and the L2 tests, as proven by the low Cronbach Alphas of 60 (Ahmed Idris), 63 (Si Pintar), 58 (Maimunah) and 65 (Ali) as in Table 11, are supported by the inconsistencies or poor discrimination of the number of test items of the Bahasa tests into the three correlation bands as presented in Table 8 (See Chapter 3.4.3).

Chapter 4

Results and Discussion of the Part I Study

4.1 Presentation and Discussion of the L1 and L2 Data.

The presentation and discussion of all the L1 and L2 data are systematically arranged in accordance with the chronology of the data analysis. It is intended that step by step evaluation and discussion of each type of data analysis will clarify the complexities involved in the analysing procedures before reaching the final conclusions. In doing so, the evaluation and discussion of the data are divided into five subheadings: Item Analysis, Factor Analysis and the Eigenvalues, Distribution of Significant Loadings, Identifying the Pattern of Relationship and Intercorrelation of Factors (Oblimin Rotation). It is believed that concurrent critical evaluation of the L1 and the L2 data under each subheading will highlight some important findings especially in terms of the strengths and the weaknesses of each set of data compared with the data from Lunzer and Gardner (1979).

4.1.1 Item Analysis

The discussion on item analysis is based on Tables 6 and 7 respectively.

Table 6: Item analysis for four Bahasa tests (n=271) Age 12

	Ahmed Idris (Grieg) Test 1		Si Pintar (Brighty) Test 2		Maimunah (Jane) Test 3		Ali (Alistair) Test 4	
<u>Subskills</u>	F *	r **	F	r	F	r	F	r
(W)	86	36	78	18	76	29	89	21
Word	52	33	95	30	82	27	38	22
Meaning	87	40	93	38	46	30	40	19
	21	15			53	23		
(WIC)	54	38	69	35	85	36	61	16
Words in	86	38	41	31	69	33	30	20
Context	46	32	95	37	74	38	91	26
	93	32	90	33	76	28	79	34
(L)	60	39	45	26	96	35	78	36
Literal	68	39	90	37	97	44	88	36
Comp.	41	28	92	27	87	29	93	29
	93	48	87	37	92	41	93	36
(ISS)	85	39	22	18	93	25	10	16
Inferences	27	27	22	31	86	38	07	31
Single			61	40			45	24
Strings			87	24			45	33
(IMS)	93	40	17	06	52	45	76	22
Inferences	35	-00	27	34	37	20	34	15
Multiple	96	32	26	36	85	46	18	14
Strings	90	37	40	27	62	41	43	18
			56	18			76	44
			34	33			38	17
(M)	04	30	26	37	76	36	17	29
Interpretation	02	13	39	46	10	14	57	39
of			24	29			30	44
Metaphor			55	24			42	28
(S)	63	54	74	48	63	39	49	45
Salient or	38	24	50	46	48	25	76	57
Main Idea								
	78	38	65	39	79	35	78	42
(J)	09	15	40	21	77	27	89	38
Forming	55	48	40	35	45	18	79	48
Judgement	43	29			32	05		
	78	35			75	43		

* % of correct responses ** correlations of item success with score on total test
(e.g. value of .77 as correlated to the perfect value of 1,
thus .77 as highly correlated), decimals omitted

A discussion on the findings for Tables 6 and 7 is based on comparing the correct responses and the correlation scores gathered and interpreted from the two tables (see the relevant discussion from page 115 to 124).

Table 7: Item Analysis For Four English Tests (n=138) Age 15

	Ahmed Idris (Grieg) Test 1		Si Pintar (Brighty) Test 2		Maimunah (Jane) Test 3		Ali (Alistair) Test 4	
<u>Subskills</u>	F *	r * *	F	r	F	r	F	r
W	73	66	75	46	91	32	61	70
	82	42	57	63	53	49	75	59
	79	40	70	62	77	54	73	32
	20	11			63	51		
WIC	54	22	54	59	54	38	67	65
	72	32	11	-29	84	36	27	29
	25	05	79	43	52	19	12	-06
L	97	-00	80	36	75	29	88	32
	91	41	50	14	99	02	77	60
	79	40	72	60	98	21	83	40
	40	43	67	54	90	51	80	54
	94	23	46	68	97	20	85	51
ISS	49	44	17	19	44	47	86	52
	59	38	47	64	88	54	12	42
			46	29			33	48
			62	48			41	36
IMS	75	64	41	21	75	63	47	41
	58	29	43	66	57	22	49	71
	88	55	51	77	83	64	42	65
	78	56	41	02	82	57	49	27
			47	26			49	32
			21	31			31	11
M	01	10	23	23	49	57	38	62
	36	57	61	67	59	57	57	80
			28	39			11	18
			70	29			51	31
S	74	48	74	74	76	45	61	63
	48	30	55	72	50	30	70	69
J	36	50	56	52	67	70	79	54
	68	69	32	32	64	38	75	36
	73	56	39	35	45	27	73	49
	40	39			42	41		
	79	44			77	64		

* % of the students with correct responses (Facility level)
* * correlations of item success with score on total test, decimals omitted

The discussion on item analysis is based on Table 6 (L1 tests) and Table 7 (L2 tests) respectively. Put simply, the function of item analysis is to provide a clearer picture as to the response or the facility level and discriminatory efficiency of every single item of the comprehension tests. These levels could guide the study in terms of the standard levels of acceptability of the test (Youngman, 1979). In their study, Lunzer and Gardner (1979) interpret the level of acceptability of the tests in terms of the percentage of correct responses (F values) and the degree of correlation between item success and score on total test (r values). Such standards are fully adopted throughout the interpretation of Table 6 and Table 7. According to Lunzer and Gardner (1979, p.49) the F values '...should not be very high or very low:...but the majority (of the F values) should fall between 20 and 80...and the more items with F values of 35-65, the better the discrimination of the test.' This implies that too easy an item would not discriminate between the good and the poor readers.

Table 6 and Table 7 show item analysis for the Bahasa (L1) and English (L2) scores. Item analysis also provides information on the correlation of every question in each skill with the total score. In a sense, items that have a high correlation with the total score '...are measuring what most of the items are measuring.' (Kline, 1994, p. 127). In this study, items with a correlation of much lower than .40 are judged as unsatisfactory. Still, too many items with a correlation above .60 are not desirable because '...the tests should also provide differential discrimination for the eight subareas.' (Lunzer and Gardner, 1979, p. 49).

Strictly speaking the figures in Tables 6 and 7 are not produced in any direct manner. The percentage of correct responses and the correlation figures have to be extracted from the correlation matrices print-out. By definition, a correlation matrix '...is a set of correlation coefficients between a number of variables.' (Kline, 1994, p. 4). In this study a huge number of correlation figures are produced from the eight variables. To simplify the figures, item analysis is used to clarify the huge correlation figures into interpretable information. The purpose of item analysis is to find the correlation of each test item with the total score of the test.

Lunzer and Gardner (1979, p. 49) say 'Items with very low or very high Fs [% of correct answers] rarely achieve a high r .' Parallel tests were conducted on the students. These would ensure the consistency of the tests. In other words, each student was subjected to four different tests. At this stage reading the index (or coefficient) of correlation entails the understanding of the value range of +1.00, perfect positive correlation and -1.00, perfect negative correlation and 0 means no correlation at all. The idea of correlation is to see whether two or more variables correspond. In this research, the two variables are the items that the students got right and the scores on the total test. Before going any further it is wise to understand that there are five data requirements that should be fulfilled before the Pearson product-moment correlation could be executed. Clegg (1990, p. 183) in his operation schedule of the correlation mentioned the following requirements:

1. Score for comparison must be paired off in some manner. Usually they will have been obtained from the same source, but this may not always be the case.
2. The relationship between the two variables must be linear one. This can be assessed from a scattergram.
3. The scores must be of at least interval status.
4. The scores must be normally distributed.
5. The two sets of scores must have similar variances.

What we can see from Table 6 and Table 7 respectively is the behaviour of the correlation of each item with the total score. How do we read the correlation coefficient (r) or the degree of agreement between the two sets of scores from the two tables? In principle, if the students do well for an individual question and if they achieve a high correlation for that question, then it will predict accurately that they will do well on the total score. Likewise, if they do badly on the individual question and they get a high correlation, then it will predict accurately whether they do badly on the total score. So, the correlation coefficient tells us more about the relationship between two variables which in this case are the individual item and the score on the total test. It is also important to note that correlation is to do with the degree of association between the two variables and

it does not imply causation. This fallacy should be avoided. Correlation can lead us to make predictions about the scores from the two variables.

Thus the coefficient, as mentioned before, could tell us that if the students do well for an individual item question and this is shown by a strong positive correlation (r) in each item of the test, then this will predict that they will also do well on the total score of the test. So, what are the coefficient values worth interpreting? Following Lunzer and Gardner's study (1979), anything that is much below .40 is not satisfactory. Furthermore, those items that have a very low or very high percentage of correct answers rarely achieve a high r . In other words, too low a percentage of correct answers for each item means the item does not discriminate between the good, average and poor test-takers; the item is too difficult for all the students. Likewise, too high a percentage of correct answers means that the item is too easy and that most of the students got it right. It is also important to note that the test items are designed to discriminate hierarchically among the eight 'subskills' and are arranged in the order of the difficulty levels. In simplifying the analysis of the correlation, the following Table 8 summarised the number of test items into three correlation bands. Lunzer and Gardner's data (1979) is used as the guiding principle in comparing the values of the correlation for all the tests done in Malaysia.

Table 8: Discrimination of the number of test items into correlation bands.

Total number of correlation items in the present study (Malaysia)		Total number of correlation items in Lunzer and Gardner (1979) (England)	
Correlation Band	Bahasa Texts (L1)	English Texts (L2)	English Texts (L1)
0.00 - 0.39	93	48	41
0.40 - 0.60	21	42	61
0.61 - 1.00	0	24	10
Total number of <i>r</i>	114	114	112

Clearly, Table 8 highlights the fact that the figures in three correlation bands for the Bahasa tests (93, 21, 0) do not display a close resemblance to the English (L2) and the Lunzer and Gardner's correlation bands. Comparatively the total number of items that are below the .40 correlation is far greater in the Bahasa texts than in the English texts. In this case 93 of the items in the Bahasa texts correlate less than .40 with the overall (total) score and practically this is poor in terms of the two variables; each test item and the total score. It is also noted that the other two bands of correlation for the Bahasa texts do not match or come close to the corresponding figures from the Lunzer and Gardner (1979) data. This low correlation of the Bahasa texts means that it is not good reliable data before factor analysis is executed. This explains why the FA data of the Bahasa texts are not as good as the English (L2) data and the Lunzer and Gardner (1979) data.

There are a number of ways of estimating the reliability of a test. Hatch and Lazaraton (1991, pp. 531-539) discuss four common methods of estimating the reliability of a test. With regard to the four methods Hatch and Lazaraton (1991, p. 531) say you should look '... (1) for consistency over time--correlation between test-retest scores; (2) for equivalence in form--correlation of parallel or comparable tests; (3) for equivalence in judgement--interrater reliability checks; and (4) for consistency *within* a test.'

In the case of this study a comparable internal consistency of the tests, as is found in Lunzer and Gardner's study (1979), is what is known as Cronbach's Alpha. It is a measure of the internal consistency and reliability of a test.

Cronbach's Alpha for the Bahasa (L1) as in Table 11 and the English (L2) tests as in Table 12 are compared with that of Lunzer and Gardner's (1979) English tests as represented in Table 3.3. The following Tables 11, 12 and 13 illustrate the values of Cronbach's Alpha and are later compared with the Cronbach values of Lunzer and Gardner's (1979) study.

Table 11: Reliability for the four Bahasa (L1) Tests(n = 271)
Pearson Correlation scores and Cronbach Alpha

Text Key

- 1 = AhmedIdris
- 2 = Si Pintar
- 3 = Maimunah
- 4 = Ali

	1	2	3	4	1+2	1+3	1+4	2+3	2+4	3+4	1+2+3+4	Cronbach Alpha
1	-	60	65	58	88	92	88	69	66	70	84	60
2	60	-	62	63	91	68	69	93	91	71	86	63
3	65	62	-	53	71	90	66	87	64	85	82	58
4	58	63	53	-	68	62	90	65	90	90	83	65
1+2	88	91	71	68	-	88	87	91	88	79	95	76
1+3	92	68	90	62	88	-	85	86	72	85	91	75
1+4	88	69	66	90	87	85	-	75	88	90	94	76
2+3	69	93	87	65	91	86	75	-	88	86	94	76
2+4	66	91	64	90	88	72	88	88	-	89	94	78
3+4	70	71	85	90	79	85	90	86	89	-	94	75
1to 4	84	86	82	83	95	91	94	94	94	94	-	86

Table 12: Reliability for the four English (L2) Tests (n = 138)
Pearson Correlation scores and Cronbach Alpha

Text Key

- 1 = AhmedIdris
- 2 = Si Pintar
- 3 = Maimunah
- 4 = Ali

	1	2	3	4	1+2	1+3	1+4	2+3	2+4	3+4	1+2+3+4	Cronbach Alpha
1	-	70	75	75	88	94	91	77	77	80	87	75
2	70	-	75	79	95	77	80	96	95	82	91	84
3	75	75	-	78	81	94	82	90	80	92	89	81
4	75	79	78	-	84	82	96	84	94	96	93	86
1+2	88	95	81	84	-	90	91	95	95	88	97	88
1+3	94	77	94	82	90	-	92	89	84	92	94	87
1+4	91	80	82	96	91	92	-	86	93	95	96	89
2+3	77	96	90	84	95	89	86	-	95	92	97	89
2+4	77	95	80	94	95	84	93	95	-	94	97	91
3+4	80	82	92	96	88	92	95	92	94	-	97	90
1to 4	87	91	89	93	97	94	96	97	97	97	-	94

Table 3.3: Reliability of the four tests

	Cronbach- Alpha	Brighty	Jane	Alistair	G+B	J+A	G+B+J+A	Edinburgh
Grieg	83	79	79	73	(94)	81	(90)	81
Brighty	83	-	81	78	(95)	85	(93)	83
Jane	85	-	-	74	85	(92)	(91)	83
Alistair	84	-	-	-	80	(94)	(90)	79
G + B	90	-	-	-	-	88	(97)	86
J + A	91	-	-	-	-	-	(97)	86
G+B+J+A	-	-	-	-	-	-	-	89

Decimals omitted from all correlation values.
n = 257 for values in the first column.
n = 212 for all the other columns.
Bracketed values are of limited worth, since they are 'contaminated'. (They represent correlations of half or quarter scores with whole scores.) Table 3.3 is taken from Lunzer and Gardner (1979, p. 52).

It is clear that the estimate of the internal reliability of the Bahasa tests is lower than the English tests (L2) and much lower than Lunzer and Gardner's internal reliability. By comparing these figures with the above Table 3.3 of Lunzer and Gardner (1979, p. 52), it is observed that the correlation values of the English (L2) tests are comparable with the English (L1) tests in UK. In this case it can be confidently stated that the internal consistency of the L2 tests is as good as the original study. There could be a possible explanation for the low internal consistency results of the Bahasa tests. There is a possibility that although the product and process of translating the four original tests into Bahasa had been thoroughly checked and much recommended by the teachers of the primary schools as very suitable to the targeted age group some other uncontrolled factors could have caused the problems. The choice of targeted words, phrases and sentences in the Bahasa texts could have some effect as to the manner of the written responses. But, in any case it is possible that translating and retranslating a text may produce different results and this could lead to incomparable results.

The idea of this research is to replicate a study within a definite time frame. Testing and retesting the comprehension tests may yield higher reliability but such action is not possible because of the time factor. Still, the Cronbach alpha figures and the teachers' recommendation are used in finding the internal consistency of the test. The high internal consistency of the L2 tests is very much expected. This is because the original English tests are not translated in any form except for the cultural meanings. In other words, the authenticity of the original texts and tests of *Grieg*, *Brighty*, *Jane* and *Alistair* in terms of vocabulary, word phrases and sentences are well kept except for some minor modifications made to suit the students' knowledge of cultural values, customs and assumptions. In a way, the background knowledge of the students is taken into consideration during the process of adapting the original texts. This adaptation is necessary in providing familiar texts and avoiding possible misunderstanding of cultural events such as the issue of colour misinterpretation as observed by McDonough (1995, p. 42):

Steffensen and Joag-Dev (1984) found that North Americans and Indians failed to identify the activity of a wedding being reported in a text because each assumed the significance of white and black was the same as in their own culture, whereas in fact they are quite different.

Cultural disorientation may be encountered by the targeted pupils if information like *Shetland Island*, *North Sea*, *Bergen* and many other proper but alien names from the *Grieg*' text are not adjusted to the local cultural setting of the Malaysian pupils. If the original texts are tested on the targeted students there is a possibility that cultural disorientation may hamper their understanding because of the mismatched cultural schemata.

Table 13 summarises the comparison of the internal reliability for all the tests in England and Malaysia. The table provides a clearer picture as to the comparative strength and weakness of the study in England and in Malaysia.

Table 13: Internal Consistencies of the Bahasa (L1) tests, English (L2) tests and Lunzer and Gardner's tests in England (Cronbach's alpha)

	Malaysia		England
	Bahasa (L1)	English (L2)	English (L1)
Test 1 (Ahmed Idris/Grieg)	60	75	83
Test 2 (Si Pintar/Brighty)	63	84	83
Test 3 Maimunah/Jane)	58	81	85
Test 4 (Ali/Alistair)	65	86	84
Decimals omitted from all correlation values. n = 271 for values in the first column. n = 138 for values in the second column. n = 257 for values in the third column.			

It must be remembered that the above reliability estimation figures in Table 13 are extracted from Tables 6 and 7 of this study and Table 3.3 of Lunzer and Gardner's study. Again, the alpha values of the L1 tests are lower than the L2 tests and far from the targeted values as found in Lunzer and Gardner's study. All the Bahasa tests have a correlation values of less than .65. The L2 texts are generally parallel in reliability to the English (L1) tests. In other words, the degree of consistency (and not perfection) of Lunzer and Gardner's tests are comparatively very much as stable as in the English (L2) tests. It is important to note that the Cronbach figures are only an estimation which '...generally only offers a lower bound for the true value. High values are desirable but Cronbach asserts that internal consistency need not be perfect for a test to be interpretable.' (Youngman, 1979, p. 180).

As mentioned earlier, it is noted that exploratory factor analysis is the chosen method of analysis used in Lunzer and Gardner's study (1979). It also involved Principal Component Analysis (PCA) and rotating the factor matrix by using Kaiser's varimax rotation in finding the correlated factors.

Principal Component Analysis (PCA)

In principle, PCA serves '...to estimate the correlation matrix and this can be done by finding the characteristic equation of the matrix.' (Kline 1994, p. 29). There are two values, called eigenvectors and eigenvalues, which are needed in making the estimation. Now, there are a few terms that need more explanation before the meaning of eigenvalues can be clearly understood.

Variance: A variance is a square of a standard deviation. In a sense, if we want to know students' placements from the mean of an examination, variance can be computed by finding the deviation of each student's score from the mean score. Summing up all the variances from the mean score should total up to zero. Then square the deviation of each score and total up the scores. This total score is known as the *sum of squares*, which is the total variability of the scores' set. The next step is to find the average of the sum of squares (the total variability score) with the number of the students. If the number of students is less than 100 then divide the score by $N-1$. This to ensure an unbiased estimate of the variance (Hatch and Lazaraton, 1991; Youngman, 1979). The concept of 'percentage of common variance' is important to this study because it explains the idea that a particular variance is being shared by some factor or factors that have emerged out of the many tests. In this study, the eight variables were factor analysed and six factors emerged with their corresponding common variance. It must be remembered that the purpose of FA is to find the factors from the many tests on the eight variables. So, FA is able to reduce the many score variances and find the common variance that has cropped up in more than one test. The reason for executing the PCA prior to the FA is that the former is able to identify all the variances and the latter is targetting the common variance.

Standard deviation (SD): The process of finding a standard deviation is just the same as the variance but it goes a little further. Instead of squaring each score from the mean, it is performed by taking the square root of the variance. It is the same as calculating the mean deviation. So, SD is a summary of the average distance of all the scores from the mean of a set of scores. SD is also associated with normal distribution; 1 SD below or above the mean, mode or median score.

As an example, 1 SD below and above the mean score of 50 with the SD score of 5 means that 68.26% of the total number of scores is between the values 45 and 55.

So, the idea of performing SD and PCA is to make some estimations for the correlation matrix. To do this requires the identification of the characteristic equation of the correlation matrix which is based on two set of values known as the characteristic vectors of the matrix (eigenvectors) and the characteristics/latent roots (eigenvalues) (Kline, 1994, p. 29). The correlation matrix (Pearson's product-moment) for the Bahasa (L1) and the English (L2) scores were computed. This initial analysis produced vast amounts of figures and it is not possible to chart the complex intercorrelations among the scores of the eight variables in this explanation. What can be done is to find the highest correlation coefficients in the matrix. By referring to Appendix B in Child (1970, p. 95), the significance level of the correlation coefficients for the Bahasa scores with a sample size of 271 is set at 0.125 at 5% level and 0.163 at 1% level. In this study the coefficients' significance level is set at 5%. For the English scores with a sample size of 138, at 5% level the coefficient is set at about 0.158. So, correlations which are above the indicated levels are worth discussion.

4.1.2. Factor Analysis (FA), Intercorrelation of Factors and the Eigenvalues.

The concepts of FA are fully discussed in chapter 3. The main objective of executing FA is to look for as many factors as possible hidden in the eight variables based on of the assumed 'subskills' tested. The many score variances from all the tests' data are extracted in order to search for the variances that are common in all the tests. The common variances reflect some underlying factors that have emerged in more than one comprehension test.

The observed factors from the factor analysis must be rotated if one wants to find interpretable factors. In general, there are two main types of rotation in factor analysis. The first one is orthogonal and it is very difficult to interpret the

factors using this method because the factors are uncorrelated or independent of one another. The second one is called rotated or oblique because the factors are correlated and this allows clearer interpretations of the factors (Kline, 1994). In this study, the oblique rotation in the *SPSS* (Statistical Package for Social Sciences) analysis package is known as *oblimin* method. This method allows three patterns of matrices known as *structure*, *pattern* and *factor correlation* matrices to be produced (Bryman and Cramer, 1990). The following Tables 14 and 15 are extracted from the *oblimin* method and most importantly they produced correlated factors for better and more rigorous interpretations of the underlying factors.

Table 14: Factor Analysis for the Four Bahasa (L1) Tests
(Oblimin Rotated Factor Matrix. Kaiser Normalization)

Tests	Subskills	Factors					
		1	2	3	4	5	6
Ahmedldris (Grieg)	W	05	08	67*	-03	-01	02
	WIC	19	-04	47*	25	-30*	03
	L	23	-34*	12	-14	-07	-28
	ISS	-03	-32*	06	-06	-32*	-26
	IMS	-14	-72*	-14	19	05	16
	M	08	11	26	45*	-07	03
	S	-01	-31*	35*	36*	09	10
	J	33*	-23	10	-15	-05	-32*
Si Pintar (Brighty)	W	-03	-26	53*	-01	02	10
	WIC	-05	03	53*	01	08	-26
	L	33*	-39*	00	-11	11	-19
	ISS	45*	08	11	-12	-05	-26
	IMS	38*	12	-17	16	07	-48*
	M	25	-11	14	16	-03	-32*
	S	34*	01	14	24	18	-14
	J	-07	09	03	02	-01	-65*
Maimunah (Jane)	W	-09	02	36*	21	24	-29
	WIC	30*	04	30*	06	03	-27
	L	06	-57*	30*	-04	01	11
	ISS	08	-47*	10	-03	13	-08
	IMS	12	-20	16	09	-12	-39*
	M	15	-26	27	-20	-48*	-04
	S	01	-15	-15	67*	05	-05
	J	-16	-34*	01	-10	14	-44*
Ali (Alistair)	W	45*	-13	11	-32*	52*	24
	WIC	-05	-17	11	03	59*	-18
	L	16	-66*	-03	09	-17	-10
	ISS	63*	06	06	03	-05	20
	IMS	59*	-15	-27	08	-05	-00
	M	55*	02	16	05	01	-06
	S	41*	-15	08	35*	06	01
	J	27	-32*	-01	22	-05	-23
% of common variance		20.6	5.1	4.5	4.0	4.0	3.8

The above figures are rounded to the nearest tens.
The * marks are the outstanding values. (Correlations of .30 and above are asterisked)

Table 15: Factor Analysis for Four English (L2) Tests
(Oblimin Rotation-Kaiser Normalization)

Test	Subskills	Factors					
		1	2	3	4	5	6
AhmedIdris (Grieg)	W	03	-04	56*	08	24	06
	WIC	18	-05	05	-49*	16	13
	L	15	09	20	-02	59*	-05
	ISS	41*	46*	04	05	41*	08
	IMS	27	-07	20	08	13	06
	S	-05	-05	04	71*	06	17
	J	25	15	06	21	13	51*
Si Pinar (Brighty)	W	51*	-21	07	10	13	14
	WIC	-09	-73*	26	11	21	-12
	L	45*	-10	02	34*	34*	-04
	ISS	35*	07	05	26	11	35*
	IMS	25	-07	04	54*	16	15
	M	57*	-06	08	-01	11	-00
	S	45*	-30*	-05	30*	18	05
	J	42*	-48*	-18	05	02	-00
Maimunah (Jane)	W	22	-22	14	-17	18	33*
	WIC	-17	05	79*	-05	04	09
	L	-16	-15	-12	05	85*	09
	ISS	51*	-08	08	-04	26	-04
	IMS	10	-15	38*	-11	53*	08
	M	80*	05	21	03	-04	-02
	S	-21	08	04	05	02	89*
	J	38*	02	08	06	38*	10
Ali (Alistair)	W	59*	-26	-06	-08	-01	25
	WIC	23	-12	70*	06	-21	-05
	L	32*	14	18	05	33*	25
	ISS	59*	03	32*	14	-07	-03
	IMS	52*	-25	14	06	01	09
	M	68*	-08	05	-08	-03	30*
	S	39*	-03	12	11	-12	50*
	J	08	-51*	02	-08	-00	52*
% of common variance		39.2	4.8	4.3	4.2	3.7	3.6
* The asterisks are the outstanding values. (Correlations of .30 and above are asterisked)							

In Table 14 and Table 15 the factor analysis figures are based on the correlation matrices computer printout. The matrices are always square and symmetric: the lower half of the matrix which is below the diagonal is the same as the upper half of the matrix. Still what is important is finding the underlying variables for each factor. What can be presumed is that if all the tests are substantially correlated in terms of the hypothetical 'subskills' then all the tests share the variance. That is, if they have common factor variance and thus if they are measuring something in common. In this case, the presumption is that if a student is good in one variable then he/she is also good in another variable across the four texts or if bad in one variable then it is also bad in another variable across the different four tests. Likewise if he/she is average in one variable then he/she is also average in another variable for all the tests.

In finding the underlying factors in Table 14 (Bahasa tests) and Table 15 (English tests) the interpretation of the factor correlation matrix is simplified in Tables 14a and 15a respectively.

Table 14a: Intercorrelation of factors for Bahasa Tests as shown in
Table 14(Factor Correlation Matrix)

Factor	1	2	3	4	5	6
1	-					
2	-23	-				
3	25	-24	-			
4	14	-06	13	-		
5	-03	-02	03	02	-	
6	-24	23	-23	-10	-02	-

Table 15a: Intercorrelation of factors for Bahasa Tests as shown in
Table 15(Factor Correlation Matrix)

Factor	1	2	3	4	5	6
1	-					
2	-23	-				
3	35	-11	-			
4	17	-08	13	-		
5	44	-14	29	12	-	
6	42	-15	27	14	36	-

It is not possible to interpret every pair of intercorrelated factors as seen in Tables 14a and 15a. But, an example of a possible interpretation is highlighted in the following discussion. It is important to note that the intercorrelation of factors can be interpreted using Table 14 and Table 15. In the simplest sense Table 14a and Table 15a are interpreted as consisting of the following elements:

In Table 14a factor 1 is negatively correlated with factor 2. In this case it is $-.23$ as observed in Table 14a and Table 15a. So, the students who do well on factor 1 actually do badly on factor 2. In Table 14, factor 1 is doing well in *Si Pintar's* and *Ali's* tests, as a pattern of high factor loadings is observed for the *L*, *ISS*, *IMS*, *M* and *S* skills for both texts. Each figure of the factor loading is actually representing a correlation of a variable with the corresponding factor, and regardless of the positive and negative signs a loading of 0.6 is regarded as high, and a loading of below 0.3 as low (Kline, 1995). The *J* skill in *Ahmed Idris's* tests emerged to be fairly correlated with factor 1. Still, the loading pattern of each subskill is in no particular hierarchical order in factor 1. For instance, the *W* skill is positively loaded (0.45) in *Ali's* text but not in the other three texts and the *WIC* skill has a high loading in *Maimunah's* text but is poorly loaded in the other 3 texts. In factor 2, the students are doing badly on every skill across the four texts. So, doing badly on factor 2 correlates negatively with doing well in factor 1. Factor 3 (Table 14) loads positively and is all about positive correlation as can be seen in the *W* and *WIC* skills across the four texts.

In factor 3, the *W* scores across the four texts are almost a representation of a crude subskills model: Vocabulary subskills versus the other subskills across all the three texts excluding *Ali's*. In other words, the pupils who got the right answers for the *W* and *WIC* questions are constantly producing the same right answers (high loadings) as observed in all the four texts. What is more, the correlation is not just about pupils who are getting the right answers, the correlation means that the pupils who get low scores on the *W* skills on *Ahmed Idris's* text also get a low score on *W* on *Si Pintar's* text. In principle, the stability of low, high or middle scores across the four passages and in the same wave of skills is important before any claim that the scores have a high loading. If those scores (the stability of high, middle or low correlation or loading scores for

the same wavelength of the assumed subskills) are shown across those four texts then it can be confidently claimed that the multiplicity of the patterned scores consists of a factor that is worthy to be extracted.

That is why it is worth interpreting factor 3 because if there are only two 'subskills', 'vocabulary skill' versus 'the other subskills', then factor 3 is a clear representation of a subskills model. If there is such a thing as a 'vocabulary subskill' it would show up by having a factor and in this case it shows up in factor 3, where in the *Ahmed Idris*, *Si Pintar* and *Maimunah* passages the same positive correlation scores are observed or came through as shown by the positive vocabulary(*W* and *WIC*) correlation scores. The pupils who do well on *W* and *WIC* in the *Ahmed Idris*, *Si Pintar* and *Maimunah* texts do not do well on *Ali's* text. In other words, the pupils who do badly on the same *W* and *WIC* skills across the same three texts do not do badly in the *Ali's* text because if they did badly the loading scores would have been higher. In a way the *W* and *WIC* loading scores from the *Ali's* text are not representing any worthy or interpretable factor. Factor 3 (Table 14) looks like a vocabulary factor. In factor 1 of Table 14, there is a pattern of comprehension subskills as shown by the high positive correlation scores of *ISS*, *IMS*, *M* and *S* in the *Si Pintar* and *Ali* texts and by the low and negative correlation scores of *ISS*, *IMS*, *M* and *S* in the *Ahmed Idris* and *Maimunah* texts.

So, what can be said is that there is a pattern of comprehension subskills in factor 1 and a likely vocabulary subskill in factor 3. The correlation of the students who got it right for the *Ahmed Idris* and *Maimunah* texts (biographical texts) with *Word (W)* and *Word in Context (WIC)* subskills is high. By the same token, both these 'subskills' are loaded low in the *Si Pintar* and *Ali* texts. In a way, there are stable patterns of either high, low or middle loadings across the four passages and these stable patterns could justify the opinion that it is a factor by itself. Clearly it can be said that factor 3 looks like a vocabulary factor. If this were the case, then by interpreting factor 1 in Table 14, a comprehension subskills pattern is observed. Factor 2 is a mixture of positive and negative figures and is not worth interpreting. Looking at factor 3 the *W* and *WIC* skills give some hints of a possible vocabulary factor. But the other

'subskills' correlations are not in any particular patterns and could only justify a unitary model. Still, regardless of the above interpretations, the factor analysis tables are not conclusive.

By referring to Table 15 (English texts) if there were such a thing as a 'word meaning' factor and the factor were supported by having a stability of either high, average or low loading scores across the four texts, it should show up in factor 3 in Table 15. What can be seen in the table is that the *W* skill in the *Ahmed Idris* text shows the high correlation score of .56 and this should show up consistently in the other three texts, but what can be observed from the table are low correlation scores for the *W* subskill of .07 in *Si Pintar's* text, .14 in *Maimunah's* and -.06 in *Ali's*. What can be seen is a variety of different scores across the four texts and these scores make no sense of interpretation. Although there is a hint of a possible vocabulary factor, such as from the average score of .56 of the *W* skill in *Ahmed Idris's* text and correspondingly a low score of .07 in *Si Pintar's* text and the high *WIC* correlation scores of .79 and .70 in *Maimunah's* and *Ali's* texts respectively, all the scores do not show a clear cut unitary model because the loading scores are not in any particular stable patterns across the four texts and across the eight subskills. This shows that factor 3 in Table 15 is not a subskill factor. It looks like a unitary comprehension factor.

It must be remembered that a factor is a group of things that go together. In other words, a pattern of consistency of the loading scores must be observed in all the texts. The pattern can be either as an individual subskill or a band or cluster of subskills across the texts. There are three possible consistency patterns that are worth interpretation: high, average and low loading scores in all the texts. In theory, underneath each loading score there are three possible groups of pupils' performances as being good, average and poor. It is important to remember that all the loading scores are figures that represent the degree of closeness with the factors. If there were such a thing as a 'Word Meaning' or a 'vocabulary' factor, it would show up as factor 3 in Table 15.

By referring to factor 1 in Table 15, all the asterisks point to a unitary skill. But before reaching any conclusion there are several assumptions that need further evaluation. In other words, in factor 1 if the *W* and the *WIC* subskills have low loadings across all the four passages, then it would suggest that factor 1 and factor 3 are vocabulary factors; a subskill pattern of comprehension is identified. In other words, two sets of subskills are identified as what should appear in factor 1 and in factor 3 respectively. But that is not the case because what can be observed is the fact that factor 1 takes up about 40 percent of the variance and there are scores that represent a unitary comprehension skill such as the high *W* scores of .51 in *Si Pintar's* text and .59 in *Ali's* text but low *W* scores in *Ahmed Idris's* and *Maimunah's* texts. This indicates that the factor is a unitary model of comprehension: The students who do well on comprehension subskills such as the *IMS* and the *J* skills also do well on *W*. Similarly, the students who do badly on *W* also do badly on comprehension subskills. The subskills model in factor 1 is not observed. In Table 15 the *ISS* and *M* subskills in factor 1 across the four texts are also worthy of interpretation. In factor 1, some positive loadings appear in the *ISS*, *IMS*, *M*, *S* and *J* scores across the four texts and they could be judged as showing a unitary model of comprehension. In other words, the low and unstable scores for the *W* and *WIC* skills across the four texts do not hamper the students from getting high loading scores for the higher order questions.

Eigenvalues:

By definition, eigenvalues are statistical figures which provide an indication of how much variance a factor is taking up. If the factor is made up on the basis of the subskills and if the eigenvalue is less than 1, the factor is judged to be contributing less variance than just one of the variables. In other words the factor is so small that it is negligible. What is wanted is an eigenvalue which is greater than one. If it is smaller than 1 it is not worth discussing. If the eigenvalue is greater than 1 then the factor might be interesting and worth interpreting. This idea is supported by Kline who says 'The larger the eigenvalue the more variance is explained by the factor.' (1994, p. 30). Refer to the following Tables 16 and 17 for the eigenvalues for both the Bahasa and the English tests.

Table 16: Bahasa (L1)

<u>Factor</u>	<u>Eigenvalue</u>	<u>Percentage of Variance</u>	<u>Cum. Percentage</u>
1	6.59	20.6	20.6
2	1.62	5.1	25.7
3	1.45	4.5	30.2
4	1.29	4.0	34.2
5	1.27	4.0	38.2
6	1.21	3.8	42.0

Table 17: English (L2)

<u>Factor</u>	<u>Eigenvalue</u>	<u>Percentage of Variance</u>	<u>Cum. Percentage</u>
1	12.54	39.2	39.2
2	1.52	4.8	43.9
3	1.36	4.3	48.2
4	1.35	4.2	52.4
5	1.19	3.7	56.1
6	1.14	3.6	59.7

Clearly, the eigenvalues for all the factors in Tables 16 and 17 are eligible for interpretation. However, only factors 1, 2 and 3 are interpreted because the bulk of the percentage of the variances are accumulated in those factors for the L1 and the L2 tests. In this study, the task of performing PCA begins with finding the correlation matrix of all the tests. Then, the next step is to find the first principal component that can explain as much as possible of

the total variability of the first data. The value given to the first component is called the eigenvalue. Then, more calculations are made to find the subsequent uncorrelated components that represent the balance of the variances from the data. What can be seen is a pattern of representation of variances across the extracted components. The new components should represent most of the information of the original data and with these components some possible factors could be extracted in the FA processes.

4.1.3 Distribution of Significant Loadings

The results of the factor analysis in Table 14 and Table 15 are discussed in a critical manner as far as the significant factors are concerned. Still, the overall hidden ingredients or skills underlying each factor are not manifested in any particular order. In order to find the underlying skills in each factor, the FA correlation scores from Table 14 and Table 15 are summarised in the following Table 18 and Table 19. The idea of summarising Table 14 and Table 15 is to find whether the many factor loadings in each factor are significant as to the eight 'subskills'. In any case, Kline (1995; 52) says:

A factor loading of 0.3 indicates that 9 per cent of the variance is accounted for by the factor. This is taken as large enough to indicate that the loading is salient. Thus in factor analyses where the sample is at least 100 subjects this is a reasonable criterion. Loadings of 0.3 or larger are regarded as significant.

In this study the adopted significant loadings of the rotated factors are the loadings above 0.30. Anything below this figure is considered worthless.

Table 18: Distribution of Significant Loadings of the Bahasa Tests.
(Summary of Table 14)

<u>Subskills</u>	<u>AhmedIdris (1)</u>	<u>Si Pintar (2)</u>	<u>Maimunah (3)</u>	<u>Ali (4)</u>
W	3	3	3	5,1,4
WIC	3,5	3	3,1	5
L	2	1,2	3,2	2
ISS	2,5	1	2	1
IMS	2	1,6	6	1
M	4	6	5	1
S	4,3	1	4	1,4
J	1,6	6	6,2	2

Table 19: Distribution of Significant Loadings for the English (L2) tests.
(Summary of Table 15)

<u>Subskills</u>	<u>AhmedIdris (1)</u>	<u>Si Pintar (2)</u>	<u>Maimunah (3)</u>	<u>Ali (4)</u>
W	3	1	6	1
WIC	4	2	3	3
L	5	1, 4, 5	5	1, 5
ISS	1, 2, 5	1, 6	1	1, 3
IMS	-	4	3, 5	1
M	1	1	1	1, 6
S	4	1, 2, 4	6	1, 6
J	6	1, 2	1, 5	2, 6

In Table 18 and Table 19 the factor loadings or the correlation scores of above 0.30 from all the four texts are taken out and are comparatively arranged and distributed in accordance with the relevant factors the loadings produced and the corresponding eight hierarchical subskills. The scattered uneven patterns of the factor loadings in both the L1 and the L2 tests are then clarified in Table 20 and Table 21.

4.1.4 Identifying the Pattern of Relationships

Table 20: Identification of factors by observing the pattern of relationship between the variables and the four Bahasa tests.

<u>Factor</u>	<u>Subskills</u>	<u>Tests</u>
1	W, WIC, L, ISS, IMS, M, S, J.	1, 2, 3, 4
2	L, ISS, IMS, J.	1, 2, 3, 4
3	W, WIC, L, S.	1, 2, 3
4	W, M, S.	1, 3, 4
5	W, WIC, ISS, M.	1, 3, 4
6	IMS, M, J.	1, 2, 3
Note: Test 1: AhmedIdris Test 2: Si Pintar Test 3: Maimunah Test 4: Ali		

In Table 20, all of the factors are loaded with different subskills and this supports the existence of unitary skills in reading comprehension. Interpreting factor 1 means that the factor is significantly loaded with various loadings with all the eight subskills in all the four tests, and the same loading pattern emerges in factor 2 albeit with only four higher order skills. In other words, vocabulary skills are not highly loaded in factor 2. There is no clear cut pattern of hierarchy of subskills in all the six factors. None of the factors consist of a salient single skill existing in isolation and this pattern of coexistence of the various subskills in each factor strongly rejects the subskills hypothesis.

Table 21: Identification of factors by observing the relationship between the variables and the four English tests (L2 texts)

<u>Factors</u>	<u>Subskills</u>	<u>Tests</u>
1	W, L, ISS, IMS, M, S, J	1, 2, 3, 4
2	WIC, ISS, S, J	1, 2, 4
3	W, WIC, ISS, IMS	1, 3, 4
4	WIC, L, IMS, S	1, 2
5	L, ISS, IMS, J	1, 2, 3, 4
6	W, ISS, M, S, J	1, 2, 3, 4
Note: Test 1: AhmedIdris Test 2: Si Pintar Test 3: Maimunah Test 4: Ali		

In both Table 20 and Table 21 the order of the subskills and the tests are specifically arranged in the order of the hierarchy of subskills and the four related tests. Factor 1 consists of most of the subskills throughout the four texts. The rest of the factors in Table 19 behave more or less the same as in Table 18. All of the six factors are clustered with various subskills and in no particular order. The mix-pattern of the eight subskills throughout the six factors supports the existence of unitary skills in reading comprehension in L2. In other words, the various significant loadings of each subskill in each text and in all the six factors as represented in Table 15 and Table 21 are interpreted as meaning that reading comprehension cannot be broken down into the eight distinct subskills and it can be concluded that reading comprehension in L2 constitutes a unitary skill. In other words, poor performance in *W* and *WIC* skills are not hampering the students' responses to higher order skills such as the *S* and the *J* skills.

4.2 Discussion of Findings for the L1 and L2 Tests.

As discussed earlier, Table 14 and Table 15 produced six factor matrices after being subjected to oblimin rotation-Kaiser normalisation with the hope that better interpretation would be possible. For an easier interpretation of the relationships among the subsets of the variables, Tables 18, 20, 19 and 21 yielded better results. A general interpretation of all the tables shows that there is no clear pattern of distinct subskills in the L1 and the L2 texts.

As argued earlier, if the data in Table 20 (L1 texts) is hierarchic in nature, then factor 1 should be in the order of the putative subskills. But it appears that the 'subskills' in factor 1 are not in any hierarchical order and far from being dominated by one single skill. What can be seen is the fact that the factor is loaded with all the subskills, which is generally comprehensive (unitary) in nature, and interestingly enough the 'inferences' skills dominated the factor in all the four tests. Clearly, factor 1 is predominantly a higher order skills factor; (*J*) judgement, (*M*) metaphorical, (*IMS*) inferences from multiple strings and (*IMS*) inferences from single strings. This phenomenon is far from the existence of lower order skills particularly the *W* skill which is not being salient or standing alone in the factor. The same behaviour exists in factor 1 in Table 21 (L2 texts).

Interpreting factor 1 in Table 15 signified the idea that the L2 respondents are at the higher level of 'meaning-making' skills. Nevertheless, it does not mean that the lower-order skills of vocabulary (*W* and *WIC*) are irrelevant or out of context. They seem to exist together with the higher order skills irrespective of the factors, particularly in factor 3 in Table 14 and Table 15. These unique loadings in factor 3 for both languages merits some discussion.

In the L1 texts, only factor 3, being vocabulary-based in nature, displays sub-skills behaviour; hierarchic or linear in nature. The same behaviour appears in the L2 texts. But, in interpreting Table 14 and Table 15, all the factors are interpreted as a whole and the behaviour of factor 3 should not overshadow the underlying meanings of the other factors. The *W* and *WIC* loadings in Table 14 and Table 15 are not highly loaded and not constant in the rest of the factors. It could be due to passage-specific-type.

It is reasonably clear that the eight subskills tested in the Bahasa (L1) and the English as a second language (L2) comprehension tests are not in any hierarchical order. There is no satisfactory evidence for the existence of separable identifiable subskills that can be built up in an hierarchical order in either language.

The results indirectly raise the question of whether the students who are poor in the so called "lower" order skills such as the 'word meaning' questions are also poor in "higher" order skills such as the 'forming judgements' questions. This question is clearly answered by the behaviour of the factor loadings in both the Bahasa and the L2 FA outputs as demonstrated in Table 14 and Table 15. The FA data suggest that regardless of the students' language ability, there are patterns of intercorrelation of subskills for each type of question tested. It is clear that the factor loadings for each subskill in L1 and L2 appear to be loaded with various subskills and in no particular hierarchical order or implicational scale: mastering the "lower" order skills is a prerequisite for the "higher" order skills.

The eight subskills identified in all the factors in both languages are clearly overlapping and far from being discrete in nature. The evidence also suggests that students who are poor in language ability should not be thought of as not being able to answer "higher order" skills. It may be that the *W* and *WIC* questions are merely testing the students' language capabilities and do not show the whole picture of their comprehension abilities because reading involves reasoning as displayed in all of the "higher" order questions. Thus, low scores in the "lower order" skills such as the *W*, *WIC* and *L* questions are not a precise indicator that the students will perform equally badly in the "higher" order questions. The FA of the L1 and L2 tests also suggests that in each factor there is a constant interplay of subskills and this implies that reading comprehension is not a lock-up procedure in which the pupils must first learn and master one subskill before learning and mastering the next subskill. The simultaneous learning of subskills is supported by Carroll (1978, p. 100) who says that:

Actually, a child can be learning a number of skills simultaneously, but will reach mastery of them at different periods in his development. From the standpoint of the teacher, this means that different skills may need to be emphasized at different periods, depending upon the characteristics of the individual child.

It is clear that the above findings support the findings of Lunzer and Gardner (1979) that reading comprehension subskills are not in any hierarchical order. Lunzer and Gardner (1979) found no conclusive evidence for the hypothesis that "...some pupils might 'possess' lower-order skills but not higher-order skills." (p. 64). This study supports the above claim and is proven by the unitary of subskills as seen in factor 1 in Table 14 and Table 15. In comparison, in Lunzer and Gardner (1979, Table 3.4 p. 56), only factor 4 showed a unitary set of subskills but there could be no clear interpretation because the loadings were high in *Brighty* and *Alistair's* tests but not in the other two tests. Lunzer and Gardner observed that the salient loadings of 'vocabulary' only appeared in factor 3 and could be attributed to a specific type of text. In a similar manner,

although 'word meaning' (*W* and *WIC*) appears to have high loadings in factor 3 in Table 14 and Table 15, the loading is not constant in all the four texts.

Interestingly enough, there is no single factor dominated by a single skill, let alone in the assumed hierarchical order. Every factor is an amalgamation of at least two subskills but of no clear interpretable pattern. There is no identifiable clear-cut factor as to the ordering of the eight skills. In other words, interpreting a factor as constituting a single construct such as 'a vocabulary factor' for factor 3 in Table 14 and Table 15 is made difficult by the existence of other subskills in each factor.

Relating this to the 'great debate' on the order in which reading skills should be taught or learned is no easy task. One could say that the patterns of the loadings in all of the factors reflect the students' simultaneous mastery of all the skills in responding and comprehending the texts. In reflection on the above simultaneous mastery of subskills, Jenkinson (in Chapman & Czerniewska, 1978) stressed the importance of "...continuing development of word recognition, the extension of knowledge of vocabulary and accurate comprehension. Several other skills must be developed simultaneously, however." (p. 186).

In conclusion, the discussed evidence does not demonstrate the hierarchical properties of subskills in reading comprehension in either the L1 and L2 comprehension tests.

Chapter 5

Conclusions of the Part I Study and its Implications for the Part II Study

Broadly speaking, the author has an impression that 'literal', 'responsive' and 'evaluative' reading are the three categories system which is represented in the hierarchical view of reading comprehension subskills. Reading skills might not be the same thing as study skills such as using a dictionary, encyclopedia or map-reading which are not in any hierarchical order compared to the three categories system. 'Literal' would be a low level skill, 'responsive' a higher skill and 'evaluative' the highest skill in hierarchical order of difficulty. Using an encyclopedia could be done at any of the three levels. So, those study skills are not necessarily in that kind of hierarchy. But an implication of the above insight is related to the concept of hierarchy in the syllabus contents of the reading curriculum of the primary and secondary schools in Malaysia. This hierarchical view should not be regarded as unproblematic. It is worth questioning the assumption noted in the Ministry document that '...All the skills listed under the various headings, are not necessarily in hierarchy. The skills and functions are stated in terms of pupil performance.' (Kementerian Pendidikan Malaysia, 1981, p. 2). Furthermore, the assumed concept of hierarchy is viewed as an age-level phenomenon.

However, in order to be fair, this study is questioning the above assumption and this is done by considering Lunzer and Gardner's (1979) work and by doing a replication study in Malaysia. It must be stressed that to be fair, the skills as quoted above are '*not necessarily*' hierarchical. Broadly speaking, it could be inferred that as a whole reading skills are assumed to be in hierarchy. The quoted sentence does not rule out the hierarchical assumption in reading skills. This assumption is proved to be inappropriate and the evidence from this study does not lend itself to such a view.

This study proves that there is no clear identifiable hierarchy of subskills in both the L1 and L2 data analysis. What can be learned is the fact that there is a trend of coexistence of different skills in each factor. The evidence does not support the hierarchical view of reading comprehension subskills. There is no evidence from the data that pupils who are poor in literal comprehension questions are unable to answer evaluative questions or vice versa. Broadly speaking, the children seem to do well on all the questions or to do badly on all of the skills. In a hierarchical view, the poor readers are thought to be good in lower order skills but not good in other higher order skills. The good readers are assumed to be good in every skill. In a sense, three reading models could be drawn up. These three models of reading subskills are fully discussed in Chapter 1. The three models are known as the unhierarchical/unrelated model, the hierarchical model and the unitary model. Furthermore, what can be observed from the FA data is that there is very little evidence of a salient factor that stands by itself and merits interpretation as representing a single subskill which if observed would mean that a linear relationship of subskills is proven--mastery of vocabulary will lead to mastery of the next difficult skill. Far from being in any hierarchical patterns the data demonstrated an interaction of skills particularly the existence of many higher-order skills such as the 'ISS' and 'IMS' in all of the factors.

Clearly it can be inferred that the students can go beyond the literal meanings with a high degree of success albeit with little correlation with vocabulary/word recognition skills. Thus, word recognition is actually '*not a total barrier*' in the quest of comprehending the texts and is far from hierarchical in nature in both the L1 and the L2 tests.

In relation to the reading process, this inconsistency in mastering word recognition is best reflected in the summary of Harrison (1992) on the related works by Smith and Goodman about a quarter of a century ago. Harrison (1992, p. 11) says:

Goodman is now thought to have been wrong in suggesting that in fluent reading only minimal

text cues are sampled in the word recognition part of the reading process. Equally, Smith is thought to have been wrong in suggesting that word recognition is 'a hindrance' to fluent reading. What we would now say is that, in fluent reading, word recognition is automatic and rapid.

Perhaps, it is the students' behaviour in their quest for meaning that resulted in 'no salient feature' as to the vocabulary skill. Although it is a common-sense view that answering a comprehension test 'vocabulary' question is likely to involve a variety of interrelated 'subskills' rather than the 'vocabulary' skill alone this view needs further investigation. Further research should not limit to 'vocabulary' test questions alone but should look into the introspective and retrospective accounts from the tested students and the students' responses on all 'subskills'. The responses could illuminate some ideas as to whether the skills are in a cumulative hierarchy or co-existed in a unitary fashion.

The findings from this study may be limited to the quantitative survey that had been carried out. It is overwhelmingly difficult to compare with other empirical findings because of the differences in materials used, social contexts, reading tasks, readers' background and the testing environment. This is what is noted by Bernhardt (1991) in her overview of vast research works conducted in Europe, Asia, Canada and the United States particularly in reading as a second language (Note: The psycholinguistic model of Goodman (1968) and Smith (1971) is the major framework reference in L2 reading research). Bernhardt (1991, p. 68) says:

The variability from study to study makes comparisons in reading development across readers, languages, and proficiency levels tantamount to impossible...the majority of second language studies are product-oriented..they rarely probe how learners *gather* those data in order to

construct understandings...second language reading studies tend to *infer* development rather than *trace* it. They infer from data gathered across different subject groups of different proficiency levels on different text types, rather than tracing it along with the development of general language proficiency within a stable passage topic domain.

Despite the difference in cultural backgrounds (U. K. and Malaysia), it can be learned that reading comprehension is empirically unitary in subskills and involves general intelligence and reasoning skills. The next challenge is to 'trace' the development of reading comprehension in both age groups and the explicit processes or strategies used by the readers in reading different text genres.

Part I of this study is concerned with the debate over hierarchical 'subskills' in reading comprehension. The collected data is strictly quantitative in nature and has no qualitative input such as through interviews or verbal discourse. In view of such limitations, a balancing perspective is needed in gaining robust, adequate and supportive qualitative evidence as to the unitary nature of the comprehension subskills. Part II of the study is an extension of the study in Part I. It traces and explores *how* the comprehension processes in both the L1 and L2 comprehension tests are manipulated within the sphere of the eight subskills.

Chapter 6

Introduction and Background to the Part II Study

6.1 Background and Rationale to the Part II Study

Part II of this study seeks to understand reading comprehension by tracing patterns of discourse, within the eight categories of 'skills', used by a small number of secondary school students in responding to comprehension texts and questions in L1 and L2.

The patterns are elicited from a series of interviews gathered from the good and the average students. The students' verbal responses in terms of why they had chosen or written a particular answer for each comprehension question were recorded and analysed by utilising the eight frames of 'skills'. In general, differences in discourse patterns are assumed to exist between the good and the average readers. The differences are also predicted in term of the languages of the texts and tests. More importantly, it is also assumed that there is a relationship between each question type and the discourse type. In other words, every single verbal response or discourse unit could be interpreted as representing one or more of the eight definitions of the 'skills'. If the relationship is true, then the eight definitions of 'skills' areas have some utility in discussing discourse patterns in reading comprehension. It is hoped that by tracing the discourse patterns of the good and the average readers, some suggestions could be advanced for understanding the comprehension processes of the readers in the L1 and L2 comprehension tests.

It is also hoped that the finding(s) could contribute to new suggestions to reading researchers and teachers in the understanding of the behavioural aspects of reading comprehension development.

Without any doubt, the data from the study in Part I lends support to the general finding of Lunzer and Gardner (1979) that individual differences in reading comprehension should be regarded as '...the pupil's ability and willingness to reflect on whatever it is he is reading.' (p. 64). But the nature of the data in Part I does not clearly identify the different abilities between the good, average and poor readers in responding to the comprehension tests. In a restricted manner all of the observed factorial patterns are subjected to inferential analysis of the three groups of respondents: good, average and poor. In Part I of this research, empirically testing the existence of subskills in reading comprehension does not provide a comprehensive view as to the processes of reasoning used by the readers in responding to the comprehension tests. It is a challenging milestone by itself and requires another dimension of data collecting for an understanding of the subskills issues from the qualitative paradigm. The findings from the data analysis on whether the eight reading comprehension subskills can be built up in any discrete hierarchical order strongly suggest the unitary nature of the skills which are not in any hierarchical pattern.

The conclusions from Part I also suggest that it is not necessary that the 'lower skills' be mastered before the 'higher skills' can be developed. Clearly, the 'subskills island' theory is rejected with a very minor exception: Vocabulary emerges in one of the six factors, albeit being far from consistent or salient in all of the factors. In a sense, vocabulary is an important ingredient in reading comprehension but lack of it is not a major block to understanding. Still, Ruddell (1994), after closely examining the literature on vocabulary acquisition and development in the context of the comprehension process, suggests that vocabulary learning is very much affected by four general factors: prior knowledge and previous experience, information available in the text, reader stance in relationship to the text and social interaction. Commenting further, Ruddell (1994) said that the *what* factors of vocabulary learning and the characteristics of a text are well studied but not the process of *how* or the application of strategies in learning words, particularly in the context of the comprehension process, and in this matter Ruddell (1994) acknowledged the importance of understanding the complex characteristics and relationships between words and the comprehension process.

What readers bring to the text is crucial in comprehension. It is important to understand the nature of the vocabulary and other skills in the context of comprehension processes. In doing so, the eight subskills are thought to represent a holistic or an ideal set of skills which are useful in the attempt to explore if there is any interactive relationship of all the skills including vocabulary. This can be studied by analysing the verbal responses of the respondents to all of the comprehension questions. Through discourse analysis we can see the similarities or differences in the comprehension answering strategies performed by the average and above average readers in responding to the L1 and L2 comprehension tests and the information is important in understanding how the readers regulate their comprehension processes.

The main objective of Part II is to explore the internal behaviour of '*the pupil's ability and willingness to reflect*' (borrowing Lunzer and Gardner's conclusion on their study) and it is done by utilising the eight subskills in the world of comprehension question answering strategies. As far as Part I is concerned, the researcher is limited to the *what* aspect of the comprehension processes. In Part II the reasoning processes or the nature of the *how* in comprehension processes are carefully studied. In other words, why an individual choose or wrote a particular answer to a particular question is exhaustively explored by means of open-ended interviews.

There are hidden puzzles surrounding the world of the 'what' findings from the FA data. If one considers the *what* findings as a collection of practised traditional skills, and one compares it with the uniqueness of the *how* data as being full of the active processes of the mind, then one could hypothesize that in trying to get a clearer understanding of reading comprehension a multiple dimensional data collecting approach could cast some light on the meaning of the term '*willingness and ability to reflect*' as used by Lunzer and Gardner (1979).

In summary, Part II of this research plans to investigate the nature of the reading comprehension answering strategies of first and second language learners in selected Malaysian secondary schools. It seeks an extension of the first research study which focused on the replication of the work of Lunzer and Gardner (1979). The first quantitative research study proved that the eight so called reading comprehension 'subskills' tested in the Bahasa and English as a second language (ESL) texts are overwhelmingly unitary in behaviour and are not in any hierarchical or uniformly discrete pattern. Other than exploring the comprehension answering strategies of the good and average readers in responding to the L1 and L2 texts, Part II of this study indirectly seeks to support the Part I findings by means of discourse analysis.

6.2. Parameters of the Study

The parameters of the study are centered on four broad aims:-

- Aim 1:** To identify patterns of relationship between the eight types of question and the coded verbal discourse units. This is done by investigating the patterns of the distribution of the coded discourse units in each type of question. The data was gathered by asking each student to write the answers for each comprehension test and immediately after completing the test, the student verbally responded as to the reasons why he/she had chosen or written the answers. Every single sentence uttered is coded within the hemisphere of the eight 'subskills'. The identified patterns of relationship are useful in enhancing our understanding of the various models of reading comprehension and in clarifying the comprehension strategies that readers use in constructing meaning when responding to the comprehension tests.

- Aim 2:** By grouping the subjects into the good and the average readers, the discourse units gained from the interviews are useful in investigating the patterns of the comprehension answering strategies employed by them.
- Aim 3:** To investigate the comprehension answering strategies of the good and the average readers in the L1 and L2 comprehension test passages and questions. By comparing the patterns of the distribution of the L1 and L2 discourse units a more precise understanding of the nature of the comprehension strategies employed by the readers in the two languages can be gained.
- Aim 4:** At the very least, to consider the implications of the results of the verbal discourse study for the work of the secondary Bahasa and the English language teachers in Malaysia, in terms of the planning of the reading comprehension curriculum, classroom teaching methodologies and selection of materials that might motivate and foster the teaching and learning of reading comprehension in the classroom.

6.3 Statement of the Problems

The product-centered research in studying reading comprehension is vast and the main methodological approaches are known as the skills approach, the taxonomy approach, the measurement approach, the factor analytic approach and the models approach. Yet, all these approaches have advanced little information in understanding the nature of reading comprehension processes (Kavale and Schreiner, 1979). The problem of this study is to find a clearer explanation of the basic reading comprehension processes involved by utilising the eight subskills

used by Lunzer and Gardner (1979). It is thought that the eight subskills are useful in categorising the verbal discourse of the comprehension questions answering strategies. In other words, the verbal responses to a particular question can be interpreted within the hemisphere of the eight subskills.

By directly investigating how the above average and average readers process the texts in their quest to make sense of the comprehension tests, some data from the processes of the strategies in answering the reading comprehension tests in L1 and L2 may be useful in advancing our understanding of these comprehension processes. With this view in mind, several research questions were investigated. The questions include the nature of the relationship between the type of question and the verbal discourse. It is worth asking if one wants to learn that in answering a comprehension question such as the *S* (finding main ideas) skill one can see a certain degree of difference in the performances of the average and above average readers by analysing the coded discourse patterns (comprehension reasoning strategies) of the respondents. All the questions are best expressed in terms of the following three hypotheses.

6.4 Hypotheses

In the light of the preceding discussion, three null hypotheses were advanced:

1- Null Hypothesis (H₀): There is no relationship in the distribution of the discourse units between the question type and the discourse type.

Alternative Hypothesis (H₁): There is a relationship in the distribution of the discourse units between the question type and the discourse type.

- 2- Null Hypothesis (Ho): There is no difference between the good and the average readers in terms of the patterns of the discourse units.
- Alternative Hypothesis (H1): There is a difference between the good and the average readers in terms of the patterns of the discourse units.
- 3- Null Hypothesis (Ho): There is no difference between the average and the good readers in terms of the distribution of the discourse units related to the language of the comprehension test passages and the language of the questions.
- Alternative Hypothesis (H1): There is a difference between the average and the good readers in terms of the distribution of the discourse units related to the language of the comprehension test passages and the language of the questions.

6.5. Importance of the Study

The decision to extend the Part I study into the area of protocol analysis was taken for a number of reasons. One aim was to extend the application of the eight 'subskills' in identifying the assumed overlapping patterns of 'subskills' through discourse analysis and strengthened the claim made in Part I that reading comprehension subskills are not in any hierarchical order and are far from being discrete. The subskills are assumed to be interactive in behaviour. The second reason was to extend our knowledge of the comprehension behaviour in the verbal reasoning used by the respondents. In a way, the self-report data gathered

from the interview could provide a more complete view of comprehension processes that were not observed from the performance scores and highlight important ideas in advancing our knowledge of the patterns of comprehension answering strategies used by the respondents.

6.6 Definitions of Terms

The following terms are frequently used throughout this study.

6.6.1 Average Students:

These students were selected upon consultation with the class teachers, the school supervisor, the headteacher and the language teachers. The mid-year language test achievements of the students were between 50 and 70 marks and none scored above 85 marks. The students' monthly academic records in both languages were also used as one of the tools in selecting the students. Their verbal communication ability was good. There were cases where the subjects had to be changed due to poor communication skills, shyness and other constraints albeit being categorised as average in language performance. These constraints could jeopardise the interviews because poor verbal skills could block the process of getting the needed information.

6.6.2 Above Average (Good) students:

As for this group, the same selection criteria as for the average readers were employed except that the mid-year language test scores were 80 and above. In other words they were categorised as very good language students in oral and written activities.

6.6.3 Comprehension Answering Strategies:

In a restricted sense, the term is used in this study to simply mean the comprehension discourse strategies of the readers in responding to the comprehension questions. The verbal inputs from the readers for each

comprehension question are critically examined and coded into the eight discourse units. Each discourse unit which is normally found to be in one complete meaningful sentence or utterance is thought to represent an embedded comprehension strategy.

Each comprehension answering strategy is rigorously coded to fit in one or more of the eight discourse types. The eight discourse types are *word meaning*, *words in context*, *literal comprehension*, *drawing inferences from a single string*, *drawing inferences from multiple strings*, *interpretation of metaphor*, *finding salient or main ideas* and *forming judgements*. In a sense, a sentence may have more than one embedded strategy and thus the coding for the sentence may have multiple discourse types. As an example, in responding to the *word meaning* questions, a student may verbally draw information from the *Word in Context*, *Literal Comprehension* and *Forming Judgement* skills.

Chapter 7

Review of Literature Pertinent to the Part II Study

Research attempts at understanding reading comprehension are not new. Other than the debatable theoretical issues of the top-down, bottom-up and interactive nature of reading comprehension (see Chapter 2), reading comprehension is now thought of as consisting of complex processes involving children's '...Oral language development, early writing experiences, encounters with environmental and other print, and myriad social interaction...' (Ruddell and Ruddell, 1994, p. 83). Thus, it is clear that children's reading comprehension is affected or at least conditioned by exposure to the above dimensions and more important is the fact that reading is no longer perceived as a thoughtless application of isolated skills or subskills by the reader. Yet, what seems to interest reading psychologists and language teachers are the complex relationships or processes of the above factors which enable children to become good readers.

The complex relationships of the factors are confirmed by reading theorists such as Goodman (1970), Smith (1973) and Miller (1963) who viewed reading as a constructive process. Still, there remains a need for clearer specification of the basic processes of how readers, particularly in the primary and secondary schools, understand the meaning of printed discourse.

Although it is impossible to know the precise process of the interpretation of meaning as used by the reader while reading a text, it is the intention of Part II of this study to explore underlying reading behaviours in the context of the eight reading comprehension subskills.

To begin with, as stated in Part I of this study, the unresolved debate of viewing reading comprehension as a unitary competence or as a set of separate identifiable subskills which are susceptible to training has raged for many years and can be traced back as far as the work of Davis (1944, 1972) and Lunzer and Gardner (1979). What can be learned from the study in Part I of this research and from the weight of contemporary research evidence is that reading

comprehension subskills seem to be a unitary competence. It is this unitary view that has stimulated the author to investigate directly how the good and the poor readers process the texts and hopefully to advance some insights into the mental behaviour behind reading comprehension processes.

In the past, product-centered research such as the taxonomy, factor analytic, skills, measurement, readability and model approaches has not advanced enough knowledge on the basic nature of reading comprehension processes. This is not surprising because this product-centered type is considered as a *post hoc* type of research simply due to the nature of the responses which were gathered after the subjects had read the texts so that the researchers had to critically infer the reading processes used by the readers. Kavale and Schreiner (1979, p. 104) consider that this type of research '... remains speculative, however, because it is once removed from the actual processes of reading comprehension.' Clearly, these researchers are far from tracing the development of the comprehension processes.

There were numerous experimental research studies in the attempt to understand the mental processes involved in reading comprehension. This includes the thinking aloud procedure of Newell and Simon (1972) where the readers had to explain their thinking in solving a problem. This study of the inner thinking or thought processes of the readers was in fact in line with that of earlier reading research by Thorndike (1917) who viewed textbook reading as requiring high level thinking, contemporarily known as reasoning processes; or that of Clark (1975, in Kavale, K and Schreiner, R., 1979) who, in addition to the view of Thorndike (1917), suggested the existence of strategies of reading in deciphering meaning from the texts.

In the past, the term 'comprehension strategies' was not fully developed by Thorndike (1917) but currently it is distinctly separated from the influential traditional definition of the behaviourists of the fifties in terms of language learning (Skinner, 1957) that encouraged drill or imitation. Dole et al. (1991, p. 242) identify a number of crucial differences between the mastery of traditional hierarchical subskills and the contemporary cognitive based views that

emphasise the interactive and constructive behaviour of readers in comprehending text. Table 22 summarizes the conceptual curriculum distinctions of the two different learning ideologies as identified by Dole et al. (1991, p. 242).

Table 22

A Summary of the Ideological Differences Between the Cognitivists' Strategic Views and Behaviourists' Traditional Subskills Views in Reading Comprehension Ability

<u>Cognitivists' Strategic Views</u>	<u>Behaviourists' Traditional Subskills Views</u>
Good readers possess flexible control of intention, adaptation and decision of which strategy to use in confronting a text.	Automatic/unconscious consistent routines applied in all texts.
Manipulating reasoning and critical thinking in making meaning (Active cognition).	Less critical in thinking and learning (less cognitive activity).
Metacognitive awareness in scaffolding the understanding.	Active practice and drill of skills may automatically be used in any reading situation.

Clearly, from Table 22, Dole et al. (1991) explicitly compare the dynamic of the conscious strategies concepts against the less conscious control of the passive traditional skills. In a sense, in strategic reading, the readers must have some intention or awareness which can help the readers to select and implement appropriate processes in creating effective understanding. This is known as 'metacognition'. Flavell (1977) identified four categories of the cognitive processes in reading: metamemory, knowledge base, basic processes and strategies. Flavell (1977) defined *strategies* as organised actions that are taken in sequence in order

to achieve the desired objective(s) and metamemory or *metacognition* as a general knowledge of the need to use any available known strategy in achieving the goal of understanding the texts.

In reviewing the literature on the definition of the term '*strategies*' in the context of reading comprehension, the author identified similarities in how the term is being defined. As an example, Kletzien (1991, p. 69) defines a *strategy* as a '... deliberate means of constructing meaning from a text when comprehension is interrupted.' After reviewing the related research literature on reading strategies Davies (1995, p. 50) proposed a definition of *strategy* as '... a physical or mental action used consciously or unconsciously with the intention of facilitating text comprehension and/or learning.' Both definitions assume an active reader interrogating the text in an attempt to maintain comprehension. Davies' (1995) work in categorizing various reading strategies across several studies merits further discussion.

After reviewing a range of nine reading research studies such as Kletzien (1991), Haarstrup (1987), Hosenfield (1977b), Sarig (1987 and ms), Scott (1990), Davies and Greene (1982) and Pritchard (1990), Davies categorised the observable and unobservable reading behaviour of the studies into five broad types of reading strategies: control reading process, monitor reading process, interact with text, utilize source of information (textual) and utilize source of information (background knowledge). Each strategy is further accompanied by many related elements. Despite the complexities in interpreting and labelling the strategies into the five classifications Davies summarises her view on reading strategies as inconclusive. She says there is '... no conclusive evidence that certain strategies are inherently more facilitating of comprehension than others.' (1995, p. 56).

Davies's remark is in line with Kletzien. Commenting on '*strategies*', Kletzien (1991, pp. 78-79) who compared the strategies used by 48 average ability students (12 of them were good comprehenders and the remaining 12 were poor comprehenders) in reading three expository texts of increasing level of difficulties, acknowledged that:

The overwhelming choices of strategies for both groups were *focusing on vocabulary, rereading previous text, making inferences, and using prior knowledge* ... It seems that the readers repeatedly utilize strategies that they feel comfortable with, and do not spontaneously try other strategies that they may know and that may be effective ..., one can conclude that the difference between the group was in *regulation*, rather than *knowledge*, of comprehension strategies.

Kletzien (1991, p. 80) also mentioned other individual variables that may affect willingness or ability to regulate strategies and one of them is categorised as '... the reader's degree of *achievement responsibility*, ... If readers feel that they have control over what happens to them in an academic situation, they will be more likely to attempt to use strategies to compensate for difficulties they encounter. On the other hand, if readers feel that their comprehension depends on the text or on the teacher, they are less likely to try to utilize any strategies that they may know.' In general, Kletzien's (1991) findings on the strategies used by the two groups seem to show that good comprehenders are more flexible and have more command of the strategies, despite the texts' difficulties, than the poor comprehenders. It seems that both Kleitzen (1991) and Davies (1995) give more credibility or emphasis to the cognitive reading comprehension strategies than to the traditional skills. This *strategies* view is supported by other cognitivists such as Duffy, Roehler, Sivan et al. (1987) and Pressley, Johnson, et al. (1989).

The above *strategies* findings, especially on how the readers were able to make use of certain reading *strategies* in the process of making sense of a particular text, may not be taught in the practice and drill of the traditional skills curriculum. In the traditional school, the repeated drill of skills which are thought to be used automatically or even unconsciously by readers may cause a setback in understanding the process of *how* the students arrived at their understanding

of the texts. It is known that the traditional evaluation procedures comprehension skills, commonly manifested in basal reader programmes, are confined to the products of the readers' reading performances that emphasize getting the right answers to the comprehension questions, and if it is viewed in that manner this traditional view of comprehension as discrete skills-thinking largely ignores or fails to explore the mental processes involved in reading comprehension especially in comprehension answering strategies.

The dynamic of viewing reading comprehension as a process rather than as product-oriented has enhanced our understanding of many ideas manifested in interchangeable related terms such as *metacognition*, *metacomprehension*, *cognitive monitoring* and *comprehension monitoring*. Baker and Brown (1984, p. 22) postulated that there is a hierarchical relationship between the terms *metacognition*, *comprehension monitoring* and *cognitive monitoring* and the relationships of the terms are stated as follows: '... metacognition, cognitive monitoring, and comprehension monitoring or metacomprehension are hierarchically related concepts. Comprehension monitoring or metacomprehension is one type of cognitive monitoring, and cognitive monitoring is a component of metacognition.' Part II of this study explores more of the concept of comprehension monitoring within the sphere of the testing of the eight categories of 'subskills'.

Dole et al. (1991) acknowledged that comprehension monitoring, a type of comprehension strategy, is important in developing comprehension. In the simplest terms comprehension monitoring is defined as '... a metacognitive process which is affected by person, strategy, and task variables.' (Wagoner (1983, p. 328). What is interesting is the fact that good readers performed better than poor readers in monitoring and regulating their comprehension (Dole et al., 1991). The poor monitoring of the ongoing comprehension by poor readers coupled with low awareness of the flexible use of comprehension strategies are also reported by Brown and Palinscar (1989). In another study Olshavsky (1976-1977) proved in her comparative study between good and poor readers that good readers (high school) used reading strategies more frequently than poor readers. But in another study on increasing passage difficulty, Olshavsky (1978)

found that both the good and the poor readers performed equally well in the number and types of reading strategies used. The different results that emerged from the various reading strategies studies may have been due to several important factors such as methods of data collection, different types of passages used, intellectual and verbal abilities of the readers and the difficulty levels of the texts given to the poor readers (Kletzien, 1991).

Due to the 'problematic' factors above, this study set out to test and interview the high school subjects who are good at verbal expression. This decision would maximise the required data which is important for the study. Should the testing and interviewing be conducted with the poor readers the chances of eliciting the desired data would be limited. This is not to suggest that poor readers are poor in verbal communication but in this study poor readers are not part of the hypotheses and research design.

In the current study, verbal report data gathered from the students is viewed as representing their thinking processes which are analysed further using the eight categories of subskills. Such 'slotting' of the verbal inputs into the eight 'subskills' is useful in understanding the comprehension answering strategies of the readers and thus could reveal pattern(s) or kind(s) of representation of reasoning strategies within the eight paradigm of 'subskills'. The thinking processes or the '*how*' in reading comprehension could provide vital information in exploring comprehension answering strategies of the readers. In fact Fyfe and Mitchell (1985, pp. 165-168) implicitly commenting on the process (how) rather than the product (what) in interview could reduce the 'by chance' phenomenon in arriving at the answer. The '*how*' in reading comprehension could cast some light for a researcher on the accuracy of the internal representation or understanding of the text (holistic meaning, tone and attitudes). The use of verbal reports in gaining '...the reasoning processes underlying higher level cognitive activity [which] ... are sometimes the only available avenue for historical or genetic analysis of mental processes.' as stated by Afflerbach and Johnston (1984, p. 308), could provide a clearer picture of the '*how*' question in comprehension.

Clearly, the *how* is much more paramount than the strategies knowledge. Preferring the term 'information skills' rather than 'advanced reading skills', Harrison (1992b, p. 327) defines the former as '... those which enable one to select, comprehend, and integrate information, usually from a number of sources.' These 'enabling' skills may be viewed as important enabling elements in understanding how readers make sense of the text. The enabling elements, which involve active interrogation of the text, require a sound mastery of text structure and Lunzer et al., (1984) after conducting a field research in developing DARTs (Directed Activity Related to Text), have identified ten common but useful and practical types of text as used across the curriculum in secondary schools (narrative, structure and mechanism, process, principle, theory, problem-solutions, historical situation, classification, instructions and theme) and *how* they should be approached; active group discussion in scaffolding the information in the text.

The essence of the DARTs project is to expose students to how different types of text operate. Examples of the pupils' tasks are; analysis (underlining, segmenting, labelling, grouping and ranking), alternative representation (listing, tabular representation, diagrammatic representation and diagram completion) and extrapolation (pupil-generated questions and imaginative extension) The impact of DARTs is that teachers must go beyond the 'skills' legacy (drill and practice) and develop the information skills. Lunzer et al. (1984 p.14) after observing children aged 10-15 and interviewing the teachers, noted that:

...most reading is reading for writing. Pupils are asked to answer questions which they read from the blackboard or from a worksheet, or they are asked to copy relevant passages in answer to a list of questions, or simply to copy a summary. In other words, reading is used as a source for revision exercises, but not as a valuable auxiliary means of instruction.

The above observation by Lunzer et al. (1984) could be interwoven with what Harrison (1992b) commented on the assertion made by the Majesty's Inspectorate that the majority of children aged 10-11 need exposure to advanced reading skills. Harrison stressed the proper execution of the strategies, a much preferred term, as a key guide to comprehension.

Comparatively, Dole et al. (1991, pp. 249-250) reviewed the process-product research where instructional acts of teachers and the products are compared between effective and less effective teachers and found several weaknesses in this research area. Despite its importance, particularly in effective classroom management such as students' time on reading tasks in developing comprehension, the weaknesses are summarised as the following:

1. It is the product of the skill-based tasks from the standardized test that differentiate the effectiveness. The *how* is not fully known.
2. The drill and practice method of instruction is judged as appropriate or valid and other methods are less known.
3. Teachers spent more 'time-on-academic-task'. It is the quantity of time spent on teaching rather than the quality of instruction.
4. Lack of multiplicity in assessing comprehension.

Dole et al. (1991, pp. 250-252) also reviewed contemporary reading instruction as explicit rather than direct. Detailed explanation of lesson content is common to both views (product vs. process instruction) but the latter is different in that:

There is no assumption that the strategy will be broken down into componential subskills. The strategy is modeled, practiced, and applied to the whole comprehension task. There is no single correct answer or a single best way to apply a particular strategy. The strategy is modeled in a variety of ways and with different tasks. There is no feed-

back about the correctness of applying a particular strategy; rather, the adaptability and flexibility of strategies are emphasized (Pearson & Dole, 1987) ... suggest that instruction, like comprehension itself, is complex and fluid.

Wallace (1992, p. 67) supported the above arguments by noting the '... generalizable strategies across languages and situations rather than specific skills.' Coming back to the lack of the *how* in the above body of research merits a move to the qualitative aspects of research. Interviewing is one of them. In assessing reading strategies, interviews could reduce the 'by-chance' phenomenon in seeking for clarification of answers. In doing so, a better accurate internal representation of text comprehension (holistic meaning, tone and attitude) could be fully achieved.

Thus, the '*how*' of the good reader - familiarity with the content, prior concepts and ideas, prior world knowledge in matching the author's assumptions in scaffolding the concepts, creating new ideas or reanalyzing his/her set of attitudes or preconceptions, combined with an understanding of the text's organisational structure due to reader's exposure to different text genres - could be assessed and shared with poor readers who might have focused on words rather than on holistic meaning. It is the ability to manipulate the internal information of the text that differentiates between the good and the poor readers. Interviews could make us aware as to how readers support their selected answer(s). There are many comprehension answering strategies that could be utilised by readers and one of them is the idea that '*reader may construct their responses from an internal representation of the content of the text.*' ((Fyfe and Mitchell, 1985, p. 165). Fyfe & Mitchell (1985, pp. 165-166) speculates:

These pupils seemed to be relying on an internal representation which was built much more on personal experience than on the contents of the text. Their responses therefore tended to be idiosyncratic. For a successful performance readers need to maintain an appropriate balance between two sources of information:

the information contained in the text, and the background knowledge and experience they use to interpret it.

It is clear that Fyfe and Mitchell (1985) suggested the importance of accurate rather than inaccurate internal representation of the content of the text in assessing performance in reading comprehension. In the context of this study there is a possibility that some readers may provide an answer which is not based on the accurate information from the text and in this regard reading teachers may be deceived by simply accepting the product without probing the rationality of the chosen answer. Interviewing the students as to the chosen answer may reveal 'the hidden agenda' of the chosen or written answer. In another study, Valencia and Pearson (1988, pp. 26-27) discussed the contemporary view of reading as a dynamic meaning making process which requires a shift to better assessment principles. After reviewing reading findings (theory, research and instruction) since the mid '70s, they outlined five suggested principles for reading teachers that embodied the interactive ('*how*') nature of active reading assessment, which seems far superior to the standardized ('*what*') test, regarding the latter as' ... inappropriate or useless for instructional decisions [since] ... teachers rely more on their hunches about children than they do on test results ... Apparently, we have been so seductively drawn to the so called objectivity, reliability, and validity of standardized norm-referenced tests that we have forgotten that they may only be minimally useful for making instructional decisions about individual students.' This notion is supported by Johnston (1983, p. 68) who claimed that much assessment practice is focused more on reliability than validity. Johnston in Pearson et al. (1984, p.175) commented on reading teachers' and researchers' biases in relying more on easy access of product results despite the acknowledged importance of process over product. Metacognitive aspects of reading, likewise, have not been stressed in reading assessment. The focus of direction should be more on individual process and contextual assessment; the '*how*' process should be researched more than the '*what*' output.

Coming back to Lunzer and Gardner's (1979 p. 64) notion of 'willingness and ability to reflect' in reading comprehension, the authors'

intention is expressed as follows:

... we are clearly not referring to some fundamental characteristic of his nervous system. What we are saying is that differences in ability and in motivation and the particular circumstances of life experience have combined to bring about a certain level of facility in interpreting the printed word which should be accepted as a point of departure for further teaching and does not stand in need of further diagnosis, especially into constituent subskills.

Thus, how students arrive at a particular interpretation of texts needs further diagnosis which merits a multidimensional approach. It is thought that the eclecticisation of two common research principles, qualitative (how) and quantitative (what), could yield better results and understanding as to the unitariness of active reading skills or strategies.

In this study, more serious attention is given to the process oriented research of reading comprehension. Such attention may increase awareness of the importance of process over product and logically any improved understanding of comprehension tests in terms of the underlying processes may be useful in guiding teachers or reading specialists in diagnosing reading comprehension problems.

Chapter 8

Design of the Part II Study

8.1 Overview of the Study

This small-scale case study seeks to understand the reading comprehension answering strategies used by good and average readers in the context of responding to L1 and L2 reading comprehension tests. It sets out to understand the comprehension answering strategies of the readers by analysing a series of face-to-face interview data. Hypothetically, it is assumed that by analysing the verbal discourses using the eight categories of subskills, the data would highlight some kind of relationship in terms of the distribution of the discourse units between the eight categorised question types and the eight discourse types. It is also predicted that there is a difference between the good and the average readers not only in terms of the patterns of the discourse units but more importantly in term of the distribution of the discourse units in both the L1 and L2 comprehension tests. It is hoped that the finding(s) would contribute some suggestions as to the pedagogical development in the reading field and advance our understanding of reading comprehension answering strategies and processes.

8.1.1. Choosing the Research Method.

The nature of this study is to understand the "how" and "why" aspects of the students' chosen or written answers to the comprehension tests. For these reasons, in this empirical primary research inquiry, a case study approach is a relevant research strategy. Yin (1994, p. 9) suggests that a case study is appropriate when:

- a "how" or "why" question is being asked about a contemporary set of events over which the investigator has little or no control.

The process of understanding "how" and "why" the students select or write a particular answer to every comprehension question asked is a challenging and time-consuming task. In the context of this study, the students were straightforwardly interviewed after completing each comprehension test. The author believed that this immediate interview strategy was the best method since the reasoning process(es) for answering each question were assumed to be fresh in their minds. During each interviewing session the author was very cautious in asking questions as to why they had chosen or written a particular answer on the grounds that any improper questioning could help or lead the students to the correct answer(s). It was thought that leading or unintentionally guiding the students to the correct answers could not yield original responses. In any case, during the face-to-face conversations, the author adopted a flexible and adaptable questioning strategy in the quest of finding out the comprehension answering strategies of the students. The interviewing methods, styles and tactics of this study were based on Robson (1993), Yin (1994) and Cohen and Manion (1989). It was also anticipated that during any interviewing session many non-verbal cues would be observed from the students and obviously they were useful in adapting to the personalities of each respondent without jeopardising the intention of the study.

Although the interviewing sessions were time-consuming the author managed to maintain a non-pressurising atmosphere. This friendly atmosphere was vital to the aim that in each session each student would provide as much information as possible (Robson, 1993). In any case, the style of the face-to-face interview was semi-structured: the author read the questions from the tests but would adjust the order of the questioning to match the context of the interviewing session (Robson, 1993).

8.1.2. Designing the Multiple Case Studies

In this study, it is acknowledged that an in-depth face-to-face single case interview would not yield adequate data and thus may not be compelling or robust enough to be regarded as a good study. Due to the nature of the above hypothetical constructs (see Chapter 6.4) a multiple-case design is of paramount

importance if the study is to produce compelling and robust findings (Yin, 1994). To be more specific, in an attempt to predict and produce the desired hypothetical results the interview procedures were being repeated or replicated for all the chosen students. It is also noted that the need for a balanced and unbiased sample in term of reading ability in both genders was also taken into account.

8.1.3 Chronology of the Part II Study

The following information outlines the chronology of events of this study:-

a) April-May 1994.

- i) Piloting the interviews in Nottingham. Testing and interviewing three Malaysian students on selected comprehension texts and tests from the eight tests taken from Part I of this research project.
- ii) Evaluating the comments and suggestions given by the students on the reading texts, comprehension tests and interview protocols.

b) May-June 1994.

- i) Designing the multiple-interviews for the chosen Malaysian schools.
- ii) A permission letter granted by the Johor Education Department to conduct the study at the selected schools.

c) July- August 1994

- i) Discussion with the headteachers and the language teachers of three selected secondary schools on the objectives of the study.
- ii) Meeting with the headteachers and class teachers of each school to identify the potential students.
- iii) Arranging the interviews with the three selected secondary schools: Maktab Sultan Abu Bakar (formerly known as English College), Sekolah Menengah (Perempuan) Sultan Ibrahim and Sekolah Menengah Sultan Ismail.
- iv) Identifying proper venues for the testing and interviewing sessions in each school.
- v) Running the comprehension tests and interviewing each student after each test had been taken. A total of 64 interviews were completed.

d) September 1994-May 1995

- i) Transcribing the 64 interviews.
- ii) Translating the transcribed data into English.
Checking the reliability of the translation.
- iii) Selecting 16 interviews for analysis.
- iv) Initial coding of the verbal discourse data using the eight definitions of 'subskills' used by Lunzer and Gardner (1979).

- v) Checking the reliability of the coding. Three co-raters were chosen and taught the objectives and the procedures of the coding.
- vi) Finding the inter-rater reliability by calculating the percentage of agreement between the author's and the raters' coding.
- vii) Proper coding of all the 16 cases.
- viii) Analysing the data using SPSS-X Release 3 statistical package.
- ix) The Mann-Whitney U - Wilcoxon Rank Sum W Test (See 8.4.4).
- x) Wilcoxon Matched-Pairs Signed-Ranks Test for two related samples (See 8.4.4).
- xi) Finding and analysing the Factor Specificity Index (FSI). (See 8.4.4).
- xii) Finding and analysing the FSI in term of the three hypothetical constructs (see chapter 6.4).
- xiii) Composing the findings. In this study the report on Part II is written in a Linear-Analytic format (Yin, 1994, p. 138).

8.2 Research Procedure 1: Pilot Case Studies.

Prior to the proper case studies a pilot project was carried out in Nottingham. The following discussion is centered on the pilot interviews and issues that emerged from the analysis of the data.

8.2.1 Piloting in Nottingham

A series of pilot interviews were carried out in Nottingham. The interviews provided ample opportunities to test the comprehension tests and to get feedback from the interviewees in terms of problems faced by them such as the readability of the L1 and L2 texts and tests. It also helped to sharpen the interviewing techniques of the researcher and to get ideas, comments and suggestions as to whether the questions posed by the researcher were proper and not directly or indirectly guiding them to any clue about the accepted answers. All of the conversations were tape-recorded and some were translated and transcribed. Two main intentions of the transcription were to identify possible loop-holes in the interviewing techniques and to spot unanticipated problems faced by the students such as the clarity of the texts and the tests.

Three Malaysian students, who were accompanying their parents studying at the University of Nottingham, aged between 13 and 15, were given the selected comprehension texts and tests. Each student was given enough time to read the texts and answer the comprehension questions. An interview was conducted immediately after each test. In total, there were more than ten visits to the students' residences. The intention of the visits was to discuss the purpose of the piloting and to familiarise the author with the students. The following table outlines the piloting programme conducted in Nottingham.

Pilot Testing in Nottingham

<u>Student Code</u>	<u>School in Nottingham</u>	<u>Age</u>	<u>Comprehension Test</u>
E. A. (Girl) (Good Reader)	C.T.C.	14	<i>Ali (L2), Maimunah (L1)</i>
R. M. A. (Girl) (Good Reader)	Glaisdale Comprehensive School, Billborough	13	<i>Ali (L2), Maimunah (L1)</i> <i>Si Pintar (L1)</i>
M. R. M. A (Boy) (Average Reader)	Glaisdale Comprehensive School, Billborough	15	<i>Ali (L2), Maimunah (L1)</i> <i>Ahmed Idris (L2)</i>

Analysing the input given by the students on the texts and the tests gave the general impression that the texts and the tests were found to be appropriate with the age levels of the students. Conducting the piloting in Nottingham was unavoidable due to the unfeasible cost and the distance factor from Nottingham to Malaysia. Thus, the author selected three Malaysian students in Nottingham who were more or less comparable to the intended Malaysian lower secondary students. These piloted students lived in Nottingham and were capable of reading and writing in Bahasa. It was assumed that if the texts and the tests were readable and testable for the matching age group in Nottingham then the chosen samples in Malaysia would be able to understand the texts and answer the comprehension questions adequately. Furthermore, the data from the FA study suggested that testing students aged 15 on the L2 texts and tests in Malaysia had not caused any problems particularly with regard to the readability of the texts and tests. This was further supported by the approval by the language teachers of the selected students prior to the tests. The texts and the tests were deemed appropriate to the reading proficiency level of the students.

8.2.2 A Report From a Selected Case

A few discussable issues emerged out of the piloted students in Nottingham. The following example illustrates some of the issues encountered during the interview session.

Student Code: E. A
Text: Maimunah (L1).
Language of test: (L1).
Discussion: In English Language.
E: Experimenter
S: Student

E: Pilih? Why did you choose ‘cari’? Is it a direct (easy) question?
Or (perhaps) it is based on your vocabulary, you know that ‘pilih’
is ‘cari’?

S: Yeah. ‘Pilih’ is..I think it is closer to ‘cari’ like ‘pilih’ is ‘to find’
which is ‘to choose’.I think.

E: All right. Aha. ‘Bakat’?

S: I’m not so sure about, about that one.

E: Yeah? According to the story, ‘bakat’ ...you said ‘seni’?

S: Emm.

E: Why not ‘keistimewaan diri’?

S: I think it has more to do with what she does like aaa...drawing.

E: You think ‘drawing’ is ‘seni’? Can you say more on that?

S: It is like emm...‘seni’ is like ‘art’ so it is art, and then she is
drawing, I think ‘bakat’ is what she does for I mean, towards that.

E: Yeah, yeah?

S: (No comments)

E: OK. Aaa..the third one? ‘peluang’?, ‘peluang’?

S: Emm...OK. This is ‘masa terbuka’. I think it is like aaa...‘opportunity’.

E: It’s like opportunity?

S: Ha.(Yes).

E: In the context of the story? (If you) base on the story?

S: Emm...aaa...It's an opportunity to learn more about art and drawing... for her.

E: Yeah? Hmm?

S: To get into the school.

E: To get into the school? All right. 'Sedar'?

S: Emm...I think that is like 'realize' to 'realize', to 'know'.

E: All right. And it is not 'hidup' or 'bangkit' or (the rest)?

S: No.

E: All right. So, number 2? 'Halaman'?

S:* I think it is just one big space.

E: One big space?

S: Yeah.

E: All right.

(A short friendly talk on her hometown in Malaysia. We shared some common past experience. Her primary school (Tengku Mariam Primary School) was next to my former Limpoon Primary School).

E: So, 'studio' is aaa...'tempat melukis gambar' (her answer), 'tempat bermain muzik', 'tempat belajar', why not 'tempat belajar'?

S:* I think 'studio' is like, musical studio is like I think is like things for recording or something recording music or whatever but I think 'studio' is for her, Maimunah, a place to draw.

E: A place to draw?

S: Yeah.

E: All right. Is that studio in the school? or where? Where the studio is?

S: I think, I don't know if it is at home. No. I don't... I think it is in school.

E: You think it is in school?

S: Yeah.

E: All right. OK. 'Tengkujuh' (monsoon)?

S:* I'm not sure about that.

E:* You aren't sure about that? Based on the story, where does the word come from? Or appear?

S: Here (pointing to the word from the text)

E: Heah. 'Musim tengkujuh'. What is 'musim' (season)?
S:* I don't know.
E: You don't know the word 'musim'?
S:* No. I can't remember.
E: You can't remember. All right. So, you choose 'nombor tujuh' (number seven)?
Why you choose 'nombor tujuh'?
S: It just sounds like it.
E: It sounds like 'tengkujuh', so you thought it is 'nombor tujuh'.
S: Yeah.
E: OK. Never mind. What about 'teduh'? Do you know the word 'teduh'?
S:* Like 'sympathy' something like that. (A wrong answer).
E: 'Teduh'?
S: Yeah, 'menacing'.
E: Ooo!
S: I can't remember.
E: 'Redup'?
S: I don't know that one (too).
E:* 'Redup' is something like aaa..when the sunlight is blocked by cloud.
Right? So, you don't get the sunlight at all, so it is called 'redup'
or 'no sunlight'...
S: Oooyeah!
E: Right?
S: Oh yeah! 'Teduh' when you are...
E:* 'Stop raining' . (She understood the idea)
S: Yeah!
E: Teduh? When the rain stopped, then it is called 'teduh'. Never mind!
Of course you know (what is meant by the word) 'hujan' right?
S: Yeah.
E: Yeah. And 'kering'? of course you know 'kering'.
S: 'Dry'.
E:* So, 'dry', so aaa...'tengkujuh' is actually 'wet season'...
S: Emm.

The verbal discourses are based on a discussion on the L1 *Maimunah* text and test. The student was an exceptionally good communicator in ESL. What is interesting is the fact that the verbal protocols are the original or primary source of information which had not been translated. The brackets are inserted by the author with the intention of clarifying the meaning of the discourses.

From the above protocols, there are several issues that need further explanation. Clearly, there is evidence of unintentional discourses, on the part of the experimenter, that may have guided the student to the proper answer (see the asterisks marked on E:*). This unintentional guidance, or discourse error, was noted after the verbal discourses had been transcribed. Once detected, it was useful and served as an interviewing guideline for the proper interviews in Malaysia. The author had to restrict himself to the intention of the study and avoid giving away any comprehension answers or clues that could lead to the answers. In a strict sense, the interviewer should only probe the nature of the comprehension answering strategies of the reader.

Apart from the answers guiding issue is the fact that the test had provided a degree of difficulty to this good reader (see the asterisks marked on S:*). The degree of difficulty for each comprehension test is due to the built-in features of the tests that are important in eliciting the comprehension answering strategies of the reader. From the above example, the student's answer for the meaning of 'musim tengkujuh' (wet season) is 'nombor tujuh' (number seven). The answer is wrong and it indicates that the student is probably guessing the answer without any reference to the text, or that her prior vocabulary knowledge does not include the exact meaning of the words. These sorts of hunches should have been further probed. In fact the author should have asked further subtle questions as to why she thought 'number seven' was the answer. Further investigating the matter is crucial in understanding the reasoning strategies of the reader. Exhaustive questioning as to why a particular answer was chosen or written is not easy. There are several factors that could have caused the experimenter to continue or to abandon further questioning of a particular answer.

Failure to probe deeply into the reasoning processes of the reader could have been due to some unrecorded factors that prompted the author to halt further questioning and proceed to the next question. These factors could have been the stress, shyness, motivation or abilities of the responder. It could also be the fact that the student would like to hide errors that are thought to be embarrassing. Non-verbal expression or body language of the students and the physical surroundings were also taken into consideration during the interviewing sessions: to ease and adapt the questioning with the intention of reducing or eliminating any sort of discomfort faced by the students and at the same time eliciting as much information as possible. The most important thing is to create a comfortable and stress-free atmosphere during the interviewing session.

In summary, unintentional guidance towards the 'correct' answers should be totally avoided. At the same time creating a comfortable stress-free interview environment for the readers to reflect as much information as possible is crucial in producing rich and reliable data. In the attempt to get rich information, the experimenter had to be aware of the problems faced by the reader, adaptable to the behaviours of the reader and creative in probing with more questions without giving a single hint as to the proper answers.

8.3 Research Procedure 2: Testing and Interviewing the 64 Cases in Malaysia

8.3.1 Selecting the Schools and the Two Groups of Students.

The selection process of the above average and the average students for the case studies are based on the predetermined design of the data collection and the permission given by the Education Department in the state of Johor to conduct the study in six selected secondary schools. From the six selected schools only three headteachers were willing to allow the author to conduct the interviews. This situation is well understood. Factors like extra-curricular activities and interference with the fixed timetables are very much anticipated. The three schools that approved the author to conduct the study are:

1. **Maktab Sultan Abu Bakar** (Sultan Abu Bakar College). Popularly and formerly known as EC (English College). It is one of the few oldest and best schools in the district of Johor Bahru, Johor. It is predominantly a boys' school. Four good readers in L1 and L2, particularly in reading comprehension, were selected at random from this school from a pool of about 300 students aged 15.
2. **Sekolah Menengah (P) Sultan Ibrahim** (Sultan Ibrahim Girls' School). Popularly known as S. I. G. S.. Like EC, it is nationally known for its superior academic and co-curriculum achievements. Four good readers, out of 280 students, were traced and interviewed. Their mid-year L1 and L2 test results, especially in reading comprehension, are all in the A category. The headteacher and the language teachers had given their total commitment in identifying the candidates.
3. **Sekolah Menengah Sultan Ismail** (Sultan Ismail Secondary School). Commonly known as S. I. This is a typical co-ed mainstream urban secondary school. This is a good school. Four boys and four girls were chosen at random from a pool of 358 students aged 15. These students, who were judged by the teachers as average comprehenders based on the L1 and L2 reading comprehension scores' report, were randomly chosen after the consultation and recommendation of the headteacher and assistant headteacher.

Again, it must be remembered that the yardstick for choosing the schools for the interviews is not the academic achievements of such schools. Rather, it is the availability and the willingness of the said schools in allowing the research to be conducted. The brief information on the three schools above should not be regarded as showing which school is academically better. It should be viewed in terms of the attempt to get the desired reading candidates for the study. It should not be inferred that Sekolah Menengah Sultan Ismail has no better reading candidates. This is far from true. In fact in any of the above schools, there are poor, average and good readers.

The crux of the matter is the proper timing of the interviews and the willingness of the students to spend their schooling hours on the tests and the interviews. It is also noted that each student had to spend between 6 to 8 hours of their schooling hours in reading, answering and verbalising the comprehension answering strategies for the four selected comprehension texts and tests. So, this time factor had to be properly arranged and executed. The researcher considered that the months of July and August were the prime time for conducting the interviews because the language teachers were expected to complete the reading syllabus before the trial Penilaian Menengah Rendah (Lower Secondary School Assessment) commonly abbreviated and called the P. M. R., which took place in mid August. The proper P. M. R. is in early October.

8.3.2 The Design of the Interviews.

The following plan outlines the 64 interviews conducted in the three selected schools.

The Plan of the Interviews.

Boy/Girl	Coding Reference	Adventurous		Biographical	
		Ali	Si Pintar	Ahmed Idris	Maimunah
Good Readers					
Girl 1	J. J.	E	M	E*	M*
Girl 2	M. M. A	E	M	M	E
Girl 3	A. M. A	E	M	E	M
Girl 4	Z. F. M. Z.	E*	M*	M	E
Boy 1	A. F. S.	M	E	M	E
Boy 2	A. H.	M	E	E*	M*
Boy 3	R. R.	M*	E*	M	E
Boy 4	I. S.	M	E	E	M
Average Readers					
Girl 1	A. Z.	E	M	E*	M*
Girl 2	N. A. M.	E	M	M	E
Girl 3	E. M.	E	M	E	M
Girl 4	L. M.	E*	M*	M	E
Boy 1	M .F. J.	M*	E*	M	E
Boy 2	M. F. S.	M	E	E*	M*
Boy 3	A. D.	M	E	M	E
Boy 4	E. S. R.	M	E	E	M
Note: 'M' stands for Malay texts and tests and 'E' for English texts and tests. The asterisk signs denote the interviews chosen for the data analysis.					

8.3.3 Collecting the Data

As stated earlier, all the 16 students were selected from three secondary schools located in the same district. Due to this same locality factor, one can raise the issue of the likelihood of discussions of the tests and answers among the selected students that may invalidate the tests. Although this problem is undesired there are several reasons why the issue of 'leaking of answers' is less likely to arise. Firstly, the researcher observed that each of the respondents live in a different area. Secondly, even if the respondents are studying in the same school or classroom, the testing and the interviewing for each student was designed in such a way that none will be able to swap information easily. As an example, the four average girl readers were tested and interviewed in the following fashion: for the *Ali* (L2) test, the girl coded as A.Z. was the first to sit the comprehension test and an hour later the girl coded as N.A.M. was called for the same test. While N.A.M. was taking the written test, the researcher interviewed A. Z. who had finished the test. By the time A.Z. had been interviewed, N.A.M. was ready for the interview and another girl, coded E.M., was called for the comprehension test. So, this cyclical routine was practised in all the schools. Finally, the students had given their assurance that they would not discuss their answers to the comprehension tests with their friends until the interviewing sessions had been completed.

Each student had to sit the four preselected comprehension tests. After each test, the student was immediately interviewed. A total of 16 interviews were chosen equally in terms of the ability of the students, the type of text, the language of the text and the gender of the students. The chosen tests and texts are asterisked as seen in the plan of the interviews (see also Appendices D (i) to D (viii) and E (i) to E (viii)).

8.4. Chronology of the Data Analysing Techniques

The techniques of analysing the data are purposely arranged from the simple overview general mean scores of the coded 7842 discourse units to the more

complex and specific comparative scores of the students' reading abilities in the L1 and L2 tests. The following information discusses the step-by-step procedures of analysing the collected data.

8.4.1 Transcribing the 64 Interviews

Transcription of the recorded conversations was carefully done by the researcher himself. In total 20 interviews out of the maximum 64 were successfully transcribed. From these 20 interviews only 16 were selected for the coding process. The initial planning was to include all 64 interviews but due to time factors the plan was altered at a later stage. During the transcribing process the researcher transcribed every single utterance recorded. Nothing was wasted on the ground that a single unrecorded utterance may make a difference in the coding.

8.4.2 Translating the Chosen 16 Interviews into English and Finding the Reliability of the Translation from Bahasa to English.

By referring to the 'Plan of the Interview' presented earlier, the asterisked texts are the 16 transcribed and translated interviews chosen for analysis. The transcribed discourse units between the author and each of the students were translated into English (see Appendix F (i) to F (iv)). Two translators were chosen to translate the verbal discourses on questions 23 and 24 of the *Si Pintar* (L2) test. The translators' main task was to transfer the meaning of each discourse unit intended by the author and the student from L1 to English. By meaning, the author is referring to the meaning intended by the author (coded as E:) and the student (coded as S:). The translators were told to be very careful in understanding the intended meaning of each discourse unit of the source language and in choosing the proper words in the target language. Each translator was given the taped verbal conversations for both questions 23 and 24, the comprehension text and the comprehension test. These were to ensure that they fully understood the context of the conversations.

The reliability of the translated conversations was checked by the researcher's supervisor and two mature Malaysian students studying for higher degrees at the University of Nottingham. The two mature students were chosen because they come from the same cultural background as the students and hence are familiar with the culture, linguistic jargon and the intended communicated meaning expressed by the researcher and the student. After a careful comparison of the translated versions by the researcher and the two translators, they claimed that the translations were accurate and reliable as far as the proper intended meaning of the author and the student were concerned.

8.4.3 Coding the Discourse Units and Finding the Reliability of the Coding.

The students' verbal responses from the 16 interviews were coded using the eight categories of 'subskills' of Lunzer and Gardner (1979, p. 44). The step-by-step tasks involved in the coding process are:

1. Reading and understanding the comprehension texts.
2. Answering all the comprehension questions.
3. Reading and understanding the eight categories of 'subskills'.
4. Reading the discourse units.
5. Understanding clearly each discourse unit in the context of the question, the text and the reader's intended meaning and coding each student's discourse unit with one or more of the eight categories of 'subskills'.
6. Writing the inferred category(ies) on the provided space at the right side of the paper. There may be more than one inferred category for each discourse unit.
7. Coding the discourse units of the students marked S:.

Before completing the coding process for all the 16 interviews, three raters were chosen to check the accuracy of the initial coding on parts of the discourse units. The three raters were given a random selection of discourse units as specified in the following format:

<u>Text</u>	<u>Student Code</u>	<u>Question Type</u>
<i>Ahmed Idris (L2)</i>	J. J.	Qs. 1a-1d (W) Qs. 2a-2c (WIC) Qs. 10 and 12 (IMS) Qs. 11 (J).
<i>Si Pintar (L2)</i>	M. F. J.	Qs. 1a-1c (W) Qs. 2a-2c (WIC) Qs. 23-24 (J)

The three raters' tasks were the same as the 7 coding steps specified in the previous page. Then, the researcher's coded discourse units were compared with the three raters'. An example of the comparison of coded discourse units is illustrated below:

Student Code: J. J.
Comprehension Test: *Ahmed Idris*
Language of Test: L2.

		CodedCategory		
		Rater	Rater	Author
		1	2	
149.	E: All right. Aaa‘What do you think Ahmed did for a living when he grew up?’ Haa can you guess what would he become when he grew up?			

			<u>CodedCategory</u>		
			Rater	Rater	Author
			1	2	
150.	S:	I think he will carry on the manufacture, the crisps..	8	4	5
151.	E:	Aha?			
152.	S:	Haa because he already had the interest in that so maybe he will carry on afterall (he) knew the recipes, mother had taught all of them maybe he will create something new!	5,8	5	5,8
153.	E:	Haa such as he invent a new recipe. whatever it is?			
154.	S:	Aha.	5,8	5	5,8
155.	E:	All right. 'There was a smell of fish in the town. A smell of fish in the town. What could be a commercial activity in the town?'			
156.	S:	Maybe manufacturing the..such as the packing of the fishes..	5,8	5,8	5,8
157.	E:	Haa?			
158.	S:	Maybe something like that.	8	5,8	5,8
159.	E:	So because of that the whole town was smelly?			
160.	S:	Something like that, maybe. (Laughing).	5,8	5,8	8
161.	E:	Aha? So, your answer is what?			
162.	S:	Aaa it's its manufacturing.	8	5,8	8
163.	E:	Aha, manufacturing of fish products, right?			
164.	S:	Aha, fish products.	8	5,8	8
165.	E:	So, it's smelly in the town?			
166.	S:	Mmm it's quite smelly.	8	-	8
167.	E:	Why smelly? What was in it? It's smelly in manufacturing, right?			
168.	S:	Maybe they used ingredients to eradicate the smell, maybe it's like we go to the pineapple's factory..	8	8	8
169.	E:	Aha?			
170.	S:	Pineapples don't really have strong smell, right?	8	8	8
171.	E:	Aha?			

172. S: But they wanted to use ingredients (chemicals) to preserve or something, the smell becomes stronger.
173. E: Aha? How did the smell could become stronger? If it's factories?
174. S: Mmm maybe they burned the or something..
175. E: There's the funnel, right? (Funnel for releasing the smell into the air)
176. S: Ahm. Right. It comes out from there.
177. E: It comes out from there? So, it's smelly?
178. S: Aha.
179. E: Aaa. 'What did Ahmed's father do for a living?'
180. S: He was a businessman.
181. E: Aaa? Is it in the story saying that he was a businessman?
182. S: Aha.

Rater	Rater	Author
1	2	
8	8	8
8	8	8
*	8	8
*	-	8
3	3	3
3	3	3

Note: The asterisk marks on the coding means that the coder inferred that only prior knowledge alone is involved and not the whole meaning of discourse type (*J*).

After the raters had completed the coding of the discourse units, the reliability of the author's coding is calculated by finding the percentage of agreement between the codings of the author and the three co-raters on each discourse unit marked S:. This was done by scoring the agreed codings as '1' and any disagreed coding as '0'. Thus, the percentage of agreement between the author and each of the three raters for the selected discourse units was calculated by totalling all the agreed codings and dividing them with the total number of the coded responses, then multiplying by 100. Thus, the formula is:

$$\text{Percentage of Agreement} = \frac{\text{Total number of '1's}}{\text{Total number of responses}} \times 100$$

After calculating the percentage of agreement for all the three raters, it was found that the individual inter-rater reliability ratings between the author and each of the three co-raters were within the range of .83 to .94. The inter-rater reliability of these three co-raters was .88. The inter-rater reliability was calculated by summing up the total number of agreements and dividing it with the total number of discourse units coded by all the three raters before multiplying the answer by 100.

8.4.4 Analysing the Coded Discourse Units for All the 16 Interviews.

All the coded discourse units were analyzed using the SPSS-X Release 3 statistical package. The analysis of the data was generally divided into three main tasks:

- a) The Mann-Whitney U - Wilcoxon Rank Sum W Test is used in determining whether there is a significant difference between the scores of two independent groups. In this case two sets of independent groups are based on gender and reading ability.
- b) Wilcoxon Matched-Pairs Signed-Ranks Test for two related samples. This test is used to find the size of the differences between two sets of related scores. The good and the average students' scores from the L1 and L2 codings are ranked and summed with the same sign.
- c) Finding and analysing the Factor Specificity Index (FSI) for all codings regardless of language, sex, ability and text-type variables. An FSI is a proportion of the number of times each factor occurs for each Question Type (QT) and Discourse Type (DT). The index is expressed in terms of the percentage of occurrence for each QT and DT. As an example, an FSI score is calculated by the following ways:

- i) First, by referring to Table 23, the FSI score for QT (*WIC*) and DT (*W*) is .26. Each student's responses that are coded as belonging to the DT (*W*) are added up and divided by the total of all the coded discourses. This gives the student's mean percentage for the DT. The same procedure is followed for all the other students. The FSI score of .26 means when responding to the QT (*WIC*) 26% of all the coded discourse units for that QT are judged as belonging to DT (*W*).
- ii) Second, all the mean scores for all eight students for the said DT are added together. This gives a total mean score.
- iii) Finally, the total mean score of all the students is divided by the number of the students. The resulting score is called the FSI. Thus the FSI of .26 is an average percentage score of all the FSI calculated separately for each of the eight students. The calculations are done using a computer program.

8.4.5 Presentation and Analysis of the Processed Discourse Data.

The presentation and analysis of the data is done in a linear fashion. Each hypothesis is debated in the context of the finding(s) from the relevant table(s). Chapter 9 discusses the results of the findings from the perspectives of the three hypothetical constructs (See Chapter 6.4).

Chapter 9

Results and Discussion of the Part II Study

For the Part II Study, following the procedures reported in Chapter 8, a corpus of data was gathered. Eight good and eight average readers, after completing four reading comprehension tests, were interviewed on the reasons why they had chosen or written their answers. Each reader was immediately interviewed after completing a comprehension test and each interview, which lasted about 50 minutes, was tape recorded. Thus, each student was subjected to four separate interviews and a total of 64 interviews yielded a total of approximately 60 hours of verbal data on the students' commentaries on the chosen and written answers to the comprehension tests.

Due to the time factor only 16 interviews, from the total of 64, were transcribed and translated into English. The chosen 16 interviews, which were taken equally from both the L1 and L2 tests, were taken from four above average and four average readers. Thus, only two interviews from each reader were analysed. The data collection for the 16 case studies was carried out as described in Chapter 8. It resulted in a corpus of 7842 discourse units, as seen in Appendices D (i) to D (viii) and E (i) to E (viii), which were carefully coded using the eight definitions of 'subskills' as used in Lunzer and Gardner's work (1979).

The reliability of the author's codings of the discourse units into the eight putative 'subskills' of Lunzer and Gardner was established in the following way: Three mature students at the University of Nottingham, all of whom were teachers with more than 5 years experience of teaching English as a second language and were fluent in English and Bahasa, were chosen for the reliability test. Two of them were Ph.D. students in Language Studies. All of them were taught how to code the discourse units. Each one then rated two sets of discourse units, taken at random from different readers (See Chapter 8.4). A total of 179 coded discourse units, chosen at random from the various question types, were then compared with the rating of the author. The inter-rater reliability of these three co-raters was .88. The individual inter-rater reliability ratings between

the author and each of the three co-raters was within the range of .83 - .94.

The intention of the coding was to explore and understand the types of reasoning processes involved in the answering strategies used by the students in responding to the L1 and L2 comprehension tests. As an example, when discussing an answer to a question on *Word Meaning (W)*, a student may talk about the vocabulary in the passage, in which case the discourse unit(s) would be coded as belonging to the Discourse Type *Word Meaning (W)*. Alternatively, in the process of explaining why he or she gave an answer, the student may talk about a judgement made about the passage as a whole. In such cases, the judgement made is coded as belonging to Discourse Type *Forming Judgement (J)*. The following two examples illustrate the latter phenomenon.

Example 1: A comprehension question on the *Word Meaning (W)* skill.

- 29. E: Mmm... 'mispronounced'?
- 30. S: Usually we associate 'mispronounced' with utterance, right? So, if it's 'wrote wrongly' it's not (the answer), in the people in the past didn't know how to write, right?

Source: See Appendix E (i)

Example 2: A comprehension question on the *Word Meaning (W)* skill.

- 1. E: Now for question 1(a), 'slinking' you choose 'creeping'. Why you choose 'creeping'?
- 2. S: This one? She went out at night, right? To find prey. She must be careful in searching her prey. She always moved slowly and quietly.
- 3. E: Aha?
- 4. S: If she is 'running', maybe could hear the footsteps around.
- 5. E: Aha?
- 6. S: 'Walking' too, (if) she's not careful enough, maybe the prey could sense them, right?
- 7. E: Aha?
- 8. S: 'Jumping' too, she's jumping too, there're preys who could see.
- 9. E: Could see?
- 10. S: She must be quiet.

Source: See Appendix D (viii)

The above examples were taken from two different students. In both cases, the students, coded as 'S', draw information which is coded as belonging to Discourse Type *Forming Judgement (J)*. In Example 1, in the process of making sense of the meaning of the word 'mispronounced', the student recalled her past experience to aid her understanding of the word. So, there is an interplay of recalling her past experience and the context of the word in the text and in this situation the responses were coded as a *Forming Judgement* discourse type. In Example 2, the student recalled his previous understanding of the animal's behaviour which is not directly stated in the text. This is clearly stated in the discourse units numbered 4, 6, 8 and 10 in Example 2. The statements in the said discourses are not directly mentioned in the text but are inferred from the text and assessed with previous experience or knowledge.

In another case, the coding of the discourse units for Question Type *Word Meaning (W)* is rather straightforward. In other words, the discourse units are coded as belonging to Discourse Type *Word Meaning (W)* only. The following scripts represents one such situation.

Example 3: Taken from Appendix E (iii).

- 25. E: What is 'threatening', actually?
- 26. S: Aa threatening (mengancam).
- 27. E: Threatening, is it?
- 28. S: Mm.
- 29. E: 'Comforting'?
- 30. S: That is comfort (selesa), probably.
- 31. E: Comfort. 'Loving'?
- 32. S: Aa that the meaning is different.
- 33. E: Different, is it?

34. S: Yes.
35. E: 'Freezing'?
36. S: It cannot be. It means cold (sejuk), right?
37. E: Aaa cold?
38. S: Mm.

In Example 3, the student's prior vocabulary knowledge is used in searching for a perfect matching of the meaning of the words. The bracketed words are the original words uttered by the student. It shows that the student knows the meaning of the L2 words and the meaning of the words are expressed in L1, as if there is an automatic interplay of understanding the meaning of words in both languages. In any case, the discourses are coded as belonging to Discourse Type *Word Meaning (M)*.

The coded discourse units were statistically analysed and are presented and discussed in two sections. Section One attempts to explore the first hypothesis that there is a relationship in the distribution of the discourse units between the Question Type (QT) and the Discourse Type (DT). This section begins with overall distribution of the mean score of each coded 'subskill' for every question type regardless of the different reading ability of the students. It seeks to understand the behaviour of the relationship between the question type and the discourse type mentioned earlier. Table 23 provides a detailed summary of the behaviour of the responses in terms of the discourse types and the question types. Also in this section, the analysis of the responses related to the question type and discourse type is developed in order to establish the argument that there is a strong relationship between question type and discourse type. Table 24 then provides a detailed breakdown of the discourse analysis for the Bahasa and English tests in order to investigate whether this relationship holds across languages.

Section Two continues to explore the remaining two hypotheses (See Chapter 6.4). It provides a detailed quantitative statistical analysis of the coded

discourse units in terms of the reading ability of the respondents in the L1 and L2 tests. Table 25 represents the general reading ability scores of the good and average readers. Tables 26 and 27 provide deeper analyses of the readers' reading abilities in terms of the two languages. Such analyses are capable of determining more precisely the nature of the scores across the question and discourse types. Differences in the scores reflect the various answering strategies used by the readers and these differences may be useful in understanding the comprehension processes of the good and the average readers.

Section One

In this section, the performance of the entire group of 16 cases, regardless of the students' reading ability and the language of the texts, is first examined.

Table 23: Proportion Scores Regardless of Reading Ability, Gender, Language and Text Types.(16 Interviews = 7842 Discourse Units)

		Discourse Type (DT)								
		W	WIC	L	ISS	IMS	M	S	J	Total
Qs. Type (QT)										
	W	.28	.29	.06	.10	.15	.00	.00	.07	.95
	WIC	.26	.29	.05	.09	.09	.00	.00	.20	.98
	L	.05	.02	.47	.18	.07	.00	.00	.17	.96
	ISS	.05	.03	.15	.52	.09	.00	.00	.16	1.0
	IMS	.01	.00	.15	.11	.36	.00	.00	.35	.98
	M	.06	.03	.09	.27	.12	.19	.00	.19	.95
	S	.03	.01	.04	.08	.48	.00	.12	.21	.97
	J	.01	.00	.01	.02	.18	.00	.00	.77	.99

Note: Each FSI score is an average score of all the FSI calculated separately for each of the eight students.

Table 23 offers an overview of the proportions of discourse units in the students' commentaries on answers related to the eight 'subskills' types. As an example, for the questions on *Word Meaning* (W) the responses of the students reflect a reliance on other 'subskills' as well and not just on *Word Meaning* per se. Each figure represents the Factor Specificity Index (FSI) for each question type. The FSI is calculated as the number of occurrences (in percentage) of each factor divided by the total number of discourse units that are interpreted and coded as belonging to that type of comprehension question. As an example, the figure .28 or 28% (row 1, column 1) represents the average percentage score of occurrences of discourse type 1 of Question Type (W) (See Chapter 8.4.4).

In Table 23, it can be seen that there is a non-random pattern of distribution of the discourse types. It may be helpful for the reader to consider the top row of data in order to get a clear view of what the table is showing. As an example, when responding to the *Word Meaning* (W) comprehension questions, the students' discourses draw upon information from other 'subskills' such as *Words in Context* (WIC) .29, *Drawing Inferences from Single Strings* (ISS) .10, *Drawing Inferences from Multiple Strings* (IMS) .15 and *Judgement* (J) .07. In this case the FSI for the (W) is 28% and the remaining 72% of the responses for that question type are unevenly distributed in other 'subskills'. So, the FSI of each QT is not entirely dominated by a single DT. The FSIs are scattered thinly or thickly throughout the eight DTs.

Reading Table 23 diagonally from the top-left to the bottom-right shows the extent to which there is an overlap between Question Type (QT) and Discourse Type (QT) as indicated by the slim rectangular shaped drawing as seen in Table 23. Initially, if we focus our attention on the high FSI scores, diagonally, then four FSI scores are worth discussing. These are the *Literal Comprehension* (L), *Drawing Inferences from Single Strings* (ISS), *Drawing Inferences from Multiple Strings* (IMS) and *Forming Judgements* (J) discourse types. It is noteworthy that when the students are talking about their understanding of the comprehension tests as a whole, they tended to talk more on the *Literal Comprehension* (L), *Drawing Inferences from Single Strings* (ISS), *Drawing Inferences from Multiple Strings* (IMS) and the *Forming Judgements* (J) discourse types. This behaviour is

supported by the high FSI scores of 47% for (*L*), 52% for (*ISS*) , 36% for (*IMS*) and 77% for the (*J*) discourse type. Again, reading diagonally, the low FSI scores for the (*M*) and (*S*) discourse types indicate that whatever the question types are, the students do not reflect much information from these 'subskills'. What these high and low scores seem to suggest is that there are certain types of discourse, in relation to the comprehension tests, that are dominating all of the discourse units.

However, while there is a variability of FSI scores within the table itself, the main point to be established is that there is a very strong relationship between Discourse Type and Question Type which is shown by the high concentration of the FSI scores along the rectangular-shaped diagonal. These four high FSI scores, the *Literal Comprehension (L)*, *Drawing Inferences from Single Strings (ISS)*, *Drawing Inferences from Multiple Strings (IMS)* and *Forming Judgements (J)* discourse types, suggest that when the students are talking about these four comprehension question types there appears to be a close relationship between Question Type and Discourse Type. By contrast, when the students are talking about vocabulary question types such as *Word Meaning (W)* and *Words in Context (WIC)*, there are smaller FSI scores as represented by the widely dispersed FSI scores seen on the first two rows of the table. Thus, the relationship between the QT and DT in the (*W*), (*WIC*), (*M*) and (*S*) skills is weaker than in the (*L*), (*ISS*), (*IMS*) and (*J*) skills.

In general, the results from Table 23 seem to suggest that regardless of the question type, the students appear to be using particular discourse types. This trend indicates that there is a strong relationship between the question types and the discourse types as confirmed by the unevenness of the distribution of the discourse units. In a sense, Table 23 does offer some support for the 'subskills' model. The table should not be thought of as reflecting cognitive subskills but it reflects a relationship between the supposed cognitive subskills and the discourse types. This is not the same as proving a link with the Factor Analysis in Part I of this study. If there is no relationship at all between the discourse type and the question type, that is if each of the comprehension question types involved all of the discourse types, then it is expected that the FSI scores should be normally distributed across the whole table. This is not the case, since the distribution of the

FSI scores is not the same for each question type. If unitary model of comprehension subskills is true then one can expect a random distribution of FSI scores in Table 23. A unitary model of comprehension subskills assumes a constant distribution of the DT scores in all the QTs (See Table 1 in Part I of this study). This is not the case either.

What is found in Table 23 is a clustering of answer types related to question types as represented by the FSI figures diagonally read from top-left to bottom-right. What is interesting is that the FSI distribution is not in any clear constant pattern and in many cases the loading of the FSI scores for DT 6 and DT 7 is zero. This pattern of instability of the FSI figures does not mean that they are unstable or arbitrary but they are reliably spread across the table thinly. As an example, in the (W) question type, the FSI loadings reflect a system of relationships among other 'subskills' and not specifically belonging to the (W) discourse type alone. What is clear is that except for the (W) *Word Meaning*, (WIC) *Words in Context*, (M) *Metaphor* and (S) *Finding Main Ideas* question types, the reasoning strategies of all the readers are very much related to the QTs as shown by the high DT scores along the diagonal. Another important point is the fact that the DTs are able to categorise the verbal reasoning strategies of all the QTs. This is shown by the high total percentage scores of at least 95% (See the column on 'Total' in Table 23).

It is clear that from the above analysis of Table 23 the strong relationship in the distribution of the discourse units between the Question Type (QT) and the Discourse Type (DT) is fully established. The results suggest that the students' verbal responses to the comprehension questions can be categorised using the eight definitions of the Discourse Types. In short the eight definitions are useful in identifying the patterns of verbal reasonings used by the students. Since the relationship between the QTs and the DTs is fully established the following discussions on the other two hypotheses are open for further critical analysis.

Table 2 4: Proportions of Discourse Units in Students' Commentaries on Answers Separated into Two Language Groups.

		Discourse Type (DT)								Total
Question Type (QT)		W	WIC	L	ISS	IMS	M	S	J	
W	B	.22*	.35*	.05	.14	.17	.00	.00	.07	1.0
	E	.35*	.24*	.06	.07	.14	.00	.00	.08	.94
WIC	B	.20	.31	.06	.09	.10	.00	.00	.24	1.0
	E	.32	.27	.03	.09	.08	.00	.00	.16	.95
L	B	.01*	.02	.60*	.15	.08	.00	.00	.10	.96
	E	.09*	.02	.33*	.21	.06	.00	.00	.24	.95
ISS	B	.02	.03	.14	.60	.07	.00	.00	.14	1.0
	E	.07	.03	.15	.43	.10	.00	.00	.18	.96
IMS	B	.00	.00	.18	.18	.34	.00	.00	.31	1.0
	E	.02	.00	.12	.05	.38	.00	.00	.40	.97
M	B	.02	.02	.11	.24	.17	.16	.00	.28	1.0
	E	.11	.05	.07	.29	.07	.23	.00	.11	.93
S	B	.01*	.00	.06	.12*	.47	.00	.12	.21	.99
	E	.05*	.01	.02	.05*	.49	.00	.12	.21	.95
J	B	.00	.00	.02	.03	.16	.00	.00	.78	.99
	E	.02	.00	.01	.02	.20	.00	.00	.76	1.0
<div>Note: B: Comprehension passages and questions in Bahasa. (8 students: 3599 discourse units) E: Comprehension passsages and questions in English. (8 students: 4243 discourse units) * P <.05 (Mann-Whitney U Wilcoxon Rank Sum W Test)</div>										

Table 24 represents a comparison of the FSI figures in the two languages of the comprehension tests. Broadly speaking, an uneven spread of FSI scores can be seen across the table regardless of the language of the tests. The table suggests that there are some differences in the discourse activities that are stable and related to the types of discourse that are produced. What is interesting is the fact that the pattern of the FSI figures for both languages is fairly familiar except in the *Word Meaning (W)* and *Words in Context (WIC)* questions. In these cases the FSI for the (W) and (WIC) DTs dominate the proportion of the coded discourses. The significant differences between the two languages for these two 'subskills' are interchangeable across the (W) and (WIC) skills. In other words, an almost identical percentages of the FSI exist for the vocabulary questions in both languages (See the horizontal FSI scores for the (W) and (WIC) skills in Table 24).

What is thought to be straightforward comprehension for (W) and (WIC) questions in L1 turns out to represent a heavy reliance on both vocabulary skills. This behaviour in L1 is rather interesting because such a heavy reliance on other skills is anticipated in the L2 tests but not in the L1 tests. When responding to the (W) QT in L1, the (ISS) and (IMS) DTs show higher FSI scores than in L2. Still, the pattern of the FSI of the (W) and (WIC) questions in both the L1 and L2 tests is a reflection of the significance of vocabulary skills dominating the entire discourse pattern of both types of question.

When reading the table diagonally (top-left to bottom-right) the FSI figures for the *Literal Comprehension (L)* 'subskills' show a clearly marked of distinction from the other 'subskills'. In the case of Bahasa, the FSI scores for the *Literal Comprehension (L)* question type marks up 60% of the discourse unit but it only accounts for 33% in English. One could argue that this statistically significant difference between the two FSI scores reflect an important difference in the reasoning behaviour of the readers. Such a phenomenon could be simply due to poorer mastery of the L2 and this is reflected by heavy reliance on the *Drawing inferences from Single Strings (ISS)* and *Forming Judgements (J)* skills. In other words, where the readers lack L2 lexical knowledge, the Discourse Types reveal additional attention by the readers to larger text units such as in the

Inferences from Single Strings (ISS) and *Forming Judgements (J)* skills. In this case, 45% of the discourse type is based on the two skills; *Inferences from Single Strings (ISS)* 21% and *Forming Judgements (J)* 24%. In contrast, in the case of the Bahasa tests, the students can respond to the *Literal Comprehension (L)* questions in a straightforward fashion without heavily relying on other skills. So, in the L2 tests, the *Literal Comprehension (L)* skill is not a 'low-level task' but requires the *Word Meaning (W)*, *Drawing Inferences from Single Strings (ISS)*, *Drawing Inferences from Multiple Strings (IMS)* and *Forming Judgements (J)* skills. The *Literal Comprehension (L)* questions in the Bahasa seems to reflect less reliance on the other Discourse Types as can be seen by the low FSI scores across the relevant row. But, in the English tests, the *Literal Comprehension* questions seem to bring up a whole array of discourse types with particularly high percentage on the *(ISS)* (21%) and *(J)* (24%) skills.

The same phenomenon can be observed in the *Drawing Inferences from Single Strings (ISS)* questions as in the *Literal Comprehension (L)* questions. The students' answering strategies or discourse types in the L2 tests for the *(ISS)* questions seem to be reflecting on the *(ISS)* less than the corresponding L1 *(ISS)* questions. The readers' responses to the L1 *(ISS)* questions appear to be much more concentrated than the L2 responses on that particular DT itself. Looking at the FSI scores for the *(ISS)* QT horizontally, the spread for the *(ISS)* questions in both languages is not much different: the responses in both languages seem to be pulling in the other six available DTs except for the *(J)* DT where the FSI score in L2 is .18 and the corresponding FSI score in L1 is .14.

Looking at the table diagonally (top-left to bottom-right), we can see the continuation of the concentration of the FSI scores on each of the remaining question type. This concentration appears in both languages. The concentration of the FSI scores in QT 6 (*Interpretation of Metaphor (M)*) is worth discussing. It seems that when students are responding to the *Interpretation of Metaphor (M)* questions, the FSI scores of L1 is slightly less than the one in L2. What is interesting is the fact the responses in L1 for that particular QT seem to rely heavily on DT *Forming Judgements (J)* (28%), DT *Drawing Inferences from Multiple Strings (IMS)* (17%) and DT *Drawing Inferences from Single Strings*

(ISS) (24%). A quite similar situation is observed when looking into the L2 responses for the same (M) QT: the responses rely more on the DT (ISS) (29%) , DT (J) (11%) and DT (W) (11%). It seems that the students' responses in L2 for QT 6 (M) use the *Word Meaning (W)* skill more than the equivalent responses in L1. It may be that the *Word Meaning (W)* skill in L2, not in L1, is important in guiding the students to respond to QT 6 (M). This may suggest that the (M) questions, such as question 17(a) ('shafts of gold') in the *Si Pintar* L2 test (see Appendix C (ii)), are posing problems to the readers. This could be due to a lack of understanding of the meaning of the word 'shafts', thus leading them to reflect more on the (W) skill. If this is the case, then there is a 'balancing-act' in making sense of the question. Such a balancing act is similar to Stanovich's (1980) interactive-compensatory model of reading comprehension. What is clear is that in either language, when responding to the QT (M) all the readers reflect considerably more on the (ISS) and (J) DTs than the remaining DTs.

In QT 7, *Finding salient or main ideas (S)*, the FSI scores of DT 7 for both languages are the same: 12% respectively. But what is interesting is that in responding to this type of question, the students' reasoning for the given answers relies almost 50% on the (IMS) skill. This happens in both languages. The students also reflect more on the use of the DT *Forming Judgements (J)*. This is shown by the 21% usage of DT (J) in both languages.

As far as QT 8, *Forming Judgements (J)*, is concerned, the bulk of responses, more than three-quarters, is concentrated on the same DT, *Forming Judgements (J)* itself: 78% of responses in L1 and 76% in L2 fall into DT *Forming Judgements (J)*. Equally important is the fact that 16% of the responses in L1 belong to the (IMS) DT, and correspondingly in L2 the figure is 20%.

In summary, Table 24 provides a clear picture as to the distribution of the discourse units between the QTs and the DTs. The distribution reflects a strong relationship between the QTs and the DTs in the two languages. This relationship is supported by the high concentration of the FSI scores diagonally and the non-random spread of the FSI scores horizontally throughout the table. More important than that is the fact that the unevenness of the distribution of the FSI

scores in both languages is direct evidence of the use of various reasoning strategies by the readers in responding to the tests. This unevenness is particularly true if one sees the marked distinction of the L1 and L2 FSI scores in DT (*L*) and DT (*ISS*). Although the FSI scores in Table 24 do not discriminate between the good and the average readers in both languages, they are remarkably useful in finding similarities and differences in the relationship of the QT and the DT. These similarities and differences are thought to be useful in guiding language teachers in understanding the basic behavioural process of comprehension answering monitoring by their students. The following section will specifically discuss further the similarities and differences in terms of the reading abilities of the good and average readers in responding to the L1 and L2 comprehension tests.

Section Two.

Table 2 5: Proportions of Discourse Units in Students' Commentaries on Answers Separated in terms of Reading Ability for both L 1 and L2 Tests(16 Interviews = 7842 Discourse Units)

		Discourse Type (DT)								Total
Qs. Type (QT)		W	WIC	L	ISS	IMS	M	S	J	
W	G	.29	.33	.07	.13	.17	.00	.00	.01	1.0
	A	.28	.26	.05	.08	.14	.00	.00	.13	.94
WIC	G	.24	.36	.05	.11	.07	.00	.00	.16	.99
	A	.29	.22	.04	.06	.11	.00	.00	.24	.96
L	G	.04	.00*	.55	.21	.06	.00	.00	.13	.99
	A	.06	.04*	.39	.15	.08	.00	.00	.20	.92
ISS	G	.04	.05	.10	.63	.04	.00	.00	.13	.99
	A	.05	.01	.19	.41	.14	.00	.00	.18	.98
IMS	G	.01	.00	.19	.10	.39	.00	.00	.31	1.0
	A	.02	.00	.11	.13	.33	.00	.00	.39	.98
M	G	.05	.02	.09	.26	.13	.29*	.00	.15	.99
	A	.08	.05	.09	.27	.11	.10*	.00	.24	.94
S	G	.01*	.01	.04	.09	.50	.00	.20	.16	1.0
	A	.05*	.00	.05	.08	.47	.00	.03	.25	.93
J	G	.00	.00	.01	.02	.22*	.00	.00	.75	1.0
	A	.02	.00	.01	.03	.14*	.00	.00	.80	1.0
<p>Note: G: Goodstudents</p> <p>A: Average students</p> <p>* P <.05 (Mann-Whitney U Wilcoxon Rank Sum W Test)</p>										

Looking at Table 25, the proportions of the discourse units in terms of reading ability for the good and the average readers represent unique patterns of response in the two groups. Again, reading the table diagonally (top-left to bottom-right), the FSI scores for the average readers are strikingly less than for the good readers except in Question Type *Forming Judgements (J)*. This single exception implies that the poorer readers need support for their answers and this is manifested by a heavy reliance on their world knowledge or *Forming Judgements (J)* DT, significantly more than the good readers, in making sense of their understanding of the test questions. This situation is further supported by the consistently higher FSI scores of the average readers than the good readers in column 8 (see the FSI figures in the vertical rectangular-shaped drawing in column 8 in Table 25); the scores are pointing in the same direction, from Question Type *Word Meaning (W)* (13%) to Question Type *Forming Judgements (J)* (80%), in terms of a heavy reliance on the *Forming Judgements (J)* skill.

The significant difference between the good and average FSI scores, taken diagonally (top-left to bottom-right) and calculated using the Wilcoxon Matched-Pairs Signed-Rank Test, is high: a two-tailed probability of .0251. Thus, the diagonal FSI differences between the good and the average students are almost unlikely due to chance alone: the diagonal FSI scores of the good readers are constantly different than the FSI scores of the average readers. The two-tailed or nondirectional hypothesis is used simply because the direction of the FSI scores differences is not predicted.

A similar pattern of slightly higher FSI scores for the average readers than for the good readers is demonstrated in column 1 as observed from Question Type *Words in Context (WIC)* (29%) to Question Type *Forming Judgements (J)* (2%). Looking from the top of column 1 (DT 1) and downwards the average readers appear to reflect more on the (W) questions than the good readers. Clearly, any assumption that there is no difference in terms of the FSI patterns between the good and the average readers is not well founded and is disproved by the above data.

What is being suggested in this chapter is that it is possible to use the discourse data to draw conclusions about the reasoning processes which occur when a reader takes a comprehension test, and what is argued is that it is in the difference between discourse types for different groups, or in different language contexts, that the stronger evidence for different reasoning processes may be found. The FSI of the DT and QT supported this argument.

In comparing the reasoning strategies used by each of the groups, significant differences were observed between the good and the average readers. Evidence of this can be found in the way the good and the average readers employed the discourse types. Both groups used a flexible approach and a variety of reasoning strategies but the good readers showed more applications of most of the discourse types, as seen diagonally (top-left to bottom-right) from the table, than the average readers. Although the average readers used similar reasoning strategies and in one particular case show a higher concentration of FSI scores (.80, as seen in Question Type *Forming Judgements (J)* on Discourse Type *Forming Judgements (J)*), they tend to activate Discourse Types less frequently than the good readers for each Question Type. It is also clear that the good readers appear to be able to focus or activate their reasoning more often on each Question Type with less reliance on other Discourse Types than the average readers. This phenomenon may hold true in a wider sense if the study is focused on the differences between good and poor readers. Again the FSI figures, as seen diagonally in the table, prove the hypothesis. With the average readers what emerges is the fact that the average readers show a heavy reliance on the Discourse Type *Forming Judgements (J)* throughout the table (see the DT (*J*) column). In a sense, the average readers tend to reflect more than the good readers on the *Forming Judgements (J)* Discourse Type. This is a type of compensatory behaviour used by the average readers in reasoning out their answers. The good readers rely less on DT (*J*) than the average readers.

Whether such consistently heavy application of the *Forming Judgements (J)* Discourse Type by the average readers throughout the interviews is proper and efficient in reasoning out the answers is subject to further study. Another point to consider in comparing the differences between the two groups is the

quality of the reasoning (i.e., being either right or wrong in their responses) in the process of making sense of their answers. This has not been a central focus in the present study, but it is important to acknowledge that it is a worthwhile issue for greater attention.

Looking closely into QT *Word Meaning (W)*, the good readers reflect slightly more than the average readers on *Words in Context (WIC)* (33%), *Literal Comprehension (L)* (7%), *Drawing Inferences from Single Strings (ISS)* (13%) and *Drawing Inferences from Multiple Strings (IMS)* (17%). The corresponding FSI scores for the average readers are (WIC) 26%, (L) 5%, (ISS) 8% and (IMS) 14%. Clearly, in the QT *Word Meaning (W)*, there is a difference between the good and the average readers in terms of using the context of the texts in facilitating their understanding and the difference is notable in the DT *Forming Judgements (J)* where the average readers use this DT 12% more than the good readers. Whether this heavy reliance on DT (J) is effective or appropriate in monitoring the average readers' comprehension is open to further investigation.

According to Stanovich's (1980, p. 63) interactive-compensatory model, '... the poor reader who has deficient word analysis skills might possibly show a *greater* reliance on contextual factors. In fact, several studies have shown this to be the case.' Although this study does not look into the patterns of the DTs of poor readers, the evidence from the data shows that both the good and the average readers interact with other skills in making sense of their answers. The various distributions of the FSI scores for QT *Word Meaning (W)* for all the good and average readers indicate that a pattern of 'compensation' does emerge but whether the compensatory acts are indications of a deficit in *Word Meaning* skills is not known.

Stanovich's (1980, p. 36) compensatory hypothesis stressed that '... a process at *any* level can compensate for deficiencies at any other level.' In this case, the 'higher-level' DTs, except for DT 6 and DT 7, are fully utilised by both groups of readers. Although the FSIs of DT 1 for QT 1 do not discriminate as to which group is poorer or weaker in the *Word Meaning (W)* skill, the indices do

indicate a greater reliance on other neighbouring skills that could provide extra sources of information. What is clear is the fact that when responding to QT 1, both groups of readers draw heavily on other skills particularly the (*W*), (*ISS*) and (*IMS*) skills and most interesting is the fact that the good readers utilise the rest of the subskills more than the average readers, except for DT (*J*). This is not to suggest that utilising more of the skills will lead to a greater precision or efficiency in the answers but what is clear is that the good readers, who are thought to be good comprehenders, display the usage of such interactive skills.

Thus, there is a possibility that the deficit in the lower-order skill (*W*) triggers the above average and average readers to reflect more on the higher-order skills such as (*ISS*), (*IMS*) and (*J*). This assumption may not be true because the term 'deficit' is not the priority of this study. But what is clear is that the FSIs reflect an active interaction of almost all the skills. This may suggest that the interactive behaviour of the skills is not compatible with either the top-down or bottom-up reading models per se. As an example, the following verbal protocols illustrate the interactive phenomenon:

Source: See Appendix D (iii). Discourse Units: 543-596

Student Code: Z.F.M.Z.

Text: Ali

Language of Text: English

543. E: OK. Next page. Number 22. Two of the following which..

544. S: '...most likely things for Ali...'..

545. E: After Ali had reached the, had been saved by the dolphin, he reached the beach right? Probably two actions from the following that he most probably would do. The most likely things he would do. First, he ran to the village looking for help, second searching for his father. Why?

546. S: I chose 'run to the village for help', at that time he had reached, right? Haa, so he what, immediately looking for help, wanting to search for his father, his father is considered dead in that sea..

547. E: Aaa?
548. S: So, he had to find the village so that, what, the people can take a boat and go out looking for it, right? Helping to find it.
549. E: Aaa. 'Look around for his father'?
550. S: Aaa there are, aaa two possibilities; he run to the village or look around for his father.
551. E: Aha?
552. S: Because usually when a child, any child, when he had arrived and he knew that his father is still there..
553. E: Aha?
554. S: He's supposed to, it's like still to look around for his father, who knows, probably his father was drifting?
555. E: Yeah, yeah. All right. If that the case, then what if he 'dry himself in the sun'?
556. S: Oh, no!
557. E: Why not?
558. S: He himself is stupid, 'dry himself in the sun.'
559. E: It is stupid?
560. S: When there's something happening over there like that, he..
561. E: There is something (which) is much more important, right?
562. S: Haa.
563. E: To save someone's life. That's what you are trying to say?
564. S: Aha.
565. E: Why not just to 'try to find a policeman'?
566. S: 'Try to find a policeman' is try to find, no!
567. E: Ha.
568. S: It's like you don't know where is the policeman, maybe it's aa because when it's in the village..
569. E: In the island, right?
570. S: Aa. This village, right? With all the inhabitants, he knows all the people with this policeman, OK..policeman, right? One policeman, and he's going to..
571. E: Haa?
572. S: Where would he go? If it were me, I would go (to find help) to..
573. E: We go to the people (because they are many), right?
574. S: Haa. Why should '...try to find...', still go and '...try to find...', if couldn't find it?

575. E: Yeah. All right. Very good. Aa 'search the beach for wreckage'?
576. S: No.
577. E: Why not?
578. S: Aa I don't think so.
579. E: It is pointless, right?
580. S: Mm. It cannot be 'go to sleep.'
581. E: 'Go to sleep', it can't be too?
582. S: Can't. No.
583. E: Why not? He's already tired?
584. S: 'Go to sleep'! I mean he's already wet..
585. E: Aa?
586. S: That after landing safely on the beach, wanted to go to the beach for a sleep?
587. E: Mm?
588. S: I mean it's really stupid.
589. E: Stupid? Or what we say that, what's the malay saying? If by doing that? He's giving priority..
590. S: To himself.
591. E: Aaa selfish, right?
592. S: Haa.
593. E: Aaa 'look for the shark'?
594. S: Just now in the middle of the sea people wanted to run away from shark, why want to 'look for the shark' when the shark is already gone and you still want to find the shark. We had saved ourselves and we want to find another trouble, what for?
595. E: Why should we [look for trouble]?
596. S: Aa [look for trouble]?
-

From the above verbal interactions, the student displayed acute awareness in reasoning out the choices of the answers. Each chosen answer was scrutinised by using prior knowledge, educated guesses and inferring the information from the text. The interplay of all the information gained helped the student to choose the answers that were thought to be appropriate.

In responding to all the standardised comprehension questions, all the readers applied complex yet interactive reasoning strategies as shown by the discussed identifiable patterns. The findings from Table 25 suggest that the good and the average readers use or activate various Discourse Types for comprehension answering strategies. The various quantified Discourse Types suggest that each group activates different patterns of reasoning processes as shown by the loading of the discourse units. The general differences of the FSI scores between the good and average readers are influenced by the QT to be answered. There are several pictures that emerge from the different patterns between the good and the average readers. They are:

1. Above average readers:

The data from the good readers suggest that there is a stronger relationship between the Question Types and Discourse Types than the average readers. The verbal reasoning responses of the good readers are highly focused on the QTs and this is proven by the higher FSI scores of the good readers' DTs than the average readers (see the diagonal FSI figures). The good readers' responses seem to be more specific and focus on every single question type itself. The higher context-specific FSI scores of the good readers, as seen diagonally, reflect the regulatory power of the readers in monitoring their comprehension. It also suggests that the comprehension discourse strategies of the readers are more active in their attempts to reason out their understanding of the questions. This may suggest that the good readers' mastery of the content of the texts make them rely less on the DT *Forming Judgement (J)*. These findings seem to indicate that the above average readers evidence a greater reflection on every QT (except on QT (J)) than the average readers. Whether such higher DT scores along the diagonal line reflects better or more successful comprehension than that of the average readers is not the focus of this study. But it can be inferred that the good readers are better than the poorer readers in monitoring their comprehension strategies (Dole et al., 1991).

2. Average readers:

In direct contrast, the data from the average readers suggest that there is a weaker relationship between the Question Types and Discourse Types, with the exception of the average readers' greater reliance on the *Forming Judgements* (J) Discourse Type than the good readers'. This greater reliance on the DT (J) may suggest that the average readers rely much more on the interplay of prior knowledge and the contents of the passages. Although this is not to say that such a strategy is a 'failing' type of strategy, it shows a kind of regulatory or monitoring strategy the readers have to adopt in making sense of the QT. Although the average readers use similar types of reasoning strategies, they are found to apply them considerably less than the above average readers in six DTs, the exceptions being DT (J) and DT (W). Once there is a weaker relationship between the DT and QT along the diagonal line, the average readers are found to utilise other DTs for each of the QTs.

The findings of this table suggest that both the above average and average readers' reasoning strategies, in all the L1 and L2 tests, reflect the usage of all the various DTs with the exception of DT 6 (M) and DT 7 (S) which are under-utilised by both groups of readers (see rows 6 and 7 in Table 25). Quantitatively, the above average readers are found to be different in their reflections of their reasoning strategies from the average readers. Whether such differences, as seen by the patterns of the FSIs, qualitatively reflect better and more effective reasoning strategies by the above average readers than by the average readers, may need further investigation.

Table 26: Proportions of Discourse Units in Students' Commentaries on
Answers Separated in terms of Reading Ability for L1
Tests(8 Interviews = 3599 Discourse Units)

		Discourse Type (DT)								Total
		W	WIC	L	ISS	IMS	M	S	J	
Qs. Type (QT)										
W	G	.20	.36	.05	.19	.20	.00	.00	.00	1.0
	A	.24	.35	.05	.09	.14	.00	.00	.13	1.0
WIC	G	.22	.42	.04	.09	.04	.00	.00	.18	.99
	A	.17	.20	.08	.09	.16	.00	.00	.30	1.0
L	G	.02	.00	.70	.17	.05	.00	.00	.09	1.0
	A	.00	.04	.53	.13	.10	.00	.00	.11	.91
ISS	G	.03	.06	.07	.65	.06	.00	.00	.12	.99
	A	.00	.00	.21	.55	.09	.00	.00	.15	1.0
IMS	G	.00	.00	.20	.17	.33	.00	.00	.31	1.0
	A	.00	.00	.16	.18	.35	.00	.00	.31	1.0
M	G	.02	.02	.11	.16	.24	.26	.00	.20	1.0
	A	.02	.01	.12	.33	.11	.06	.00	.36	1.0
S	G	.00	.01	.04	.13	.45	.00	.20	.16	.99
	A	.01	.00	.07	.10	.50	.00	.04	.25	.97
J	G	.00	.00	.02	.03	.19	.00	.00	.76	1.0
	A	.00	.00	.02	.03	.14	.00	.00	.80	.99
<u>Note</u>		G: Good students (above average students)								
		A: Average students								

Table 26 provides a comparison between the comprehension answering strategies of the above average and average students when responding to the L1 tests. It seeks to find out if there is a difference between the two groups of students in terms of their monitoring of their reasoning strategies from the perspective of the L1 test data. It must be remembered that there were only four good and four average readers tested in the comprehension tests and the FSI scores are not tested for the significance difference. This is unavoidable due to the small number of students for each group. Each of the four Bahasa texts and tests was only tested once in each group of readers. In fact each of the FSI scores in Table 25 is derived from the FSI scores in Tables 26 and 27. Due to the subdivision of Table 25 into Tables 26 and 27, the significance differences in the later tables are thought to be weak. Although the discussions for Table 25 have proven the fact that there is a difference between the good and the average readers in terms of the patterns of the discourse units, the following discussions for Tables 26 and 27 attempt to probe deeper into the same hypothesis in the context of the L1 and L2 texts and tests separately. To begin with, there are several findings gained from Table 26 that merit considerable discussion.

First, by looking diagonally (top-left to bottom-right), a different pattern emerges in the two groups of readers. The good students respond more in discourse types (*WIC*) (42%), (*L*) (70%), (*ISS*) (65%), (*M*) (26%) and (*S*) (20%) and slightly less in discourse types (*W*) (20%), (*IMS*) (33%) and (*J*) (76%) than the average students. In other words, the average students' FSI scores are slightly higher in percentages than the good students in (*W*) (24%), (*IMS*) (35%) and (*J*) (80%) and considerably less than the good readers in the (*M*) (6%) and (*S*) (4%) skills. All the good and average students' FSI scores are high in their utilisation of the (*L*), (*ISS*) and (*J*) skills. The overall significance difference of all of the good and average FSI scores taken diagonally (top-left to bottom-right) and calculated using the Wilcoxon Matched-Pairs Signed-Rank Test is very low: a two-tailed probability of .0929.

This low probability is not surprising due to the possibility that in general the average readers reading comprehension answering strategies are more or less the same as the good readers. Another possible factor of such overall low

probability could be due to the language of the text and test: probably, the average readers were able to respond to the L1 texts and tests as good as the good readers. In any case, the overall diagonal significance difference of the good and the average students' FSI scores could be enhanced by increasing the number of cases. In this case, a total of only four Bahasa tests were subjected for the interviews for each group of students. Still, despite the insignificance of the overall diagonal FSI scores between the good and average students, the main issue is to explore the reading comprehension answering strategies of the two groups when responding to L1 texts and tests. The two-tailed or nondirectional hypothesis is used simply because the direction of the differences between the good and the average readers' FSI scores could not be predicted.

The remarkable differences of the FSI scores between the two groups, as seen along the diagonal line, are noted in the (*WIC*), (*L*), (*ISS*), (*M*) and (*S*) skills where the good students utilise these skills more often than the average students. Again, for all the students, being higher or lower in the FSI scores is not a direct indication of being deficient in any particular skill knowledge. It is just a reflection of their patterns of thinking and patterns of interaction with the information gained either from the texts, tests and prior knowledge when responding to the tests. But, as in Table 25, the above average readers are superior to the average readers in concentrating and regulating their thought to the five types of questions. This trait of question-specific concentration of thinking is less in the average readers. These remarkable differences in the ability or perhaps willingness to regulate the skills are open to many other factors such as motivation and cognition of the students (Kletzien, 1991). It can be said that differences in the FSI patterns between the two groups are a clear indication that the above average comprehenders are more focused than the average comprehenders in the stated five DTs in constructing the reasoning of the tests. This focus-oriented phenomenon may well suggest that the above-average students are more active in their regulation of their reasoning strategies.

Second, by looking downwards from the top of column 8, there is a remarkable difference between the two groups. The average students are utilising more of the (*J*) skills than the above average readers in all QTs except in QT

(*IMS*) where the FSI scores for the DT (*J*) of both groups are tied at 31%. In other words, the 'rich' reflection of the average readers on the (*J*) skill is an indication of the use of compensatory-argumentation in making sense of the questions. The good readers' reliance on such skill is far less than the average readers except when responding to the specific (*J*) QT itself. In fact in QT 1, the FSI score of the DT (*J*) of the above average readers is 0%. This indicates that the above average readers do not resort at all to the (*J*) skill in responding to the (*W*) questions. What is clear is that the average readers repeatedly utilise more of the (*J*) skill than the above average readers. Whether such a strategy is effective is not known.

Third, columns 3, 4 and 5 reflect heavy utilisation of the (*L*), (*ISS*) and (*IMS*) skills by both groups. Other than the (*J*) skills in column 8, these three skills are loaded with the various concentration of the FSI scores. In these three columns the FSI scores of the good and average readers are patternised with higher or lower FSI scores.

Fourth, looking horizontally at QT 1 and QT 2, two interesting patterns emerge. When responding to the (*W*) questions, all the readers almost equally regulate their reasoning within the QT itself, but consistently reflecting on the use of the (*WIC*) skill: the FSI scores of the (*WIC*) skill for both the groups are around a third of their reasoning strategies (see the FSI scores of DT (*WIC*) for QT (*W*)). When responding to the QT (*WIC*), the FSI scores of DT (*WIC*) of the good readers are about 100% more than the average readers. For the same QT (*WIC*), the good readers are utilising the DT (*W*) skill 5% more than the average readers.

Finally, looking down the DT in columns 1 and 2 for the (*L*), (*ISS*), (*IMS*), (*M*), (*S*) and (*J*) QTs suggest that these skills are thinly reflected by both groups of readers. This probably reflects the fact that (*W*) and (*WIC*) skills are not causing problems to all the students and thus more reflection can be focused on the other skills. If these two 'low-order' skills were causing problems then it could be predicted that the loading of the FSI scores for these two skills should be high in the first two columns. But that is not the case.

In summary, the above average readers were found to be able to concentrate more of their reasoning processes within the boundary of the QT than the average readers (see the diagonal FSI). For each QT, both the groups are found to be relying on other DTs which are not specific to the QT itself. This degree of 'reliance factor' is not always the same for both groups. Sometimes the above average readers are reflecting more than the average readers and vice versa. So, the reliance factors are not constant for the two groups with the exception of DT (*J*): the average readers reflect more on the (*J*) skill than the good readers.

There is no doubt that the good readers exhibit a stronger relationship between the Question Types (QT) and Discourse Types (DT) than the average readers in (*WIC*), (*L*), (*ISS*), (*M*), and (*S*) skills. The good readers show a slightly weaker relationship between the QT and DT than the average readers in (*WM*), (*IMS*) and (*J*) skills. But even though the average readers are weaker than the good readers in the said relationships, all the FSI scores of the average readers, as seen diagonally from the table, suggest that they are able to focus their thinking in most of the DTs except in DT (*M*) and DT (*S*). The superior relationship of QT and DT in the above average readers suggests that the focus of their reasoning processes are very much tied to the specific question tested.

The findings of this table suggest that both the above average and average readers' reasoning strategies, in all the tests, reflect the usage of all the various DTs with the exception of DT (*M*) and DT (*S*) which are remarkably less reflected on by the average readers (see the specific DTs and QTs of the (*M*) and (*S*) skills in Table 26). Other than the specific (*W*) and (*WIC*) questions (see the high FSI scores for DTs (*W*) and (*WIC*)) both type of readers are inferred to be competent in understanding the words encountered in the texts, and thus vocabulary was not causing difficulties for them. Otherwise the FSI scores for DTs 1 and 2 for all the QTs would have been high. But this is not the case. More importantly, although the tests were in their native language, the above average readers were found to be more active in their reasoning processes than the average readers. This conclusion is clearly supported by the diagonal distribution of the FSI scores of the above average readers.

Table 27: Proportions of Discourse Units in Students' Commentaries on Answers Separated in terms of Reading Ability for L2 Tests (8 Interviews = 4243 Discourse Units)

		Discourse Type (DT)								
		W	WIC	L	ISS	IMS	M	S	J	Total
Qs. Type (QT)										
W	G	.38	.30	.08	.07	.13	.00	.00	.03	.99
	A	.32	.17	.05	.06	.14	.00	.00	.13	.87
WIC	G	.25	.29	.07	.13	.09	.00	.00	.13	.96
	A	.40	.25	.00	.04	.07	.00	.00	.19	.95
L	G	.06	.00	.42	.25	.07	.00	.00	.18	.98
	A	.11	.03	.25	.17	.05	.00	.00	.30	.91
ISS	G	.04	.05	.12	.60	.03	.00	.00	.14	.98
	A	.10	.01	.18	.26	.18	.00	.00	.21	.94
IMS	G	.01	.00	.18	.02	.45	.00	.00	.32	.98
	A	.03	.00	.06	.08	.31	.00	.00	.48	.96
M	G	.07	.03	.07	.36	.02	.32	.00	.11	.98
	A	.14	.08	.07	.21	.12	.14	.00	.12	.88
S	G	.01	.01	.03	.04	.54	.00	.21	.16	1.0
	A	.09	.01	.02	.06	.44	.01	.03	.26	.92
J	G	.00	.00	.00	.01	.25	.00	.00	.73	.99
	A	.03	.00	.01	.02	.14	.00	.00	.80	1.0
Note: G: Good students										
A: Average students										

In Table 27, the reading ability of the two groups of students are more or less the same as in Table 26. Again, as in Table 26, all the FSI significance differences of the good and average readers are not calculated due to the small sample size. But there are striking differences between the two groups especially in the context of the L2 reading comprehension tests.

First, looking diagonally (top-left to bottom-right), the above average readers are found to be consistently higher in the FSI scores than the average readers in all of the skills except in the (*J*) skill, where the average readers are slightly higher by 7%. But when comparing the diagonal FSI scores of all the readers in Table 27 with the diagonal FSI scores in Table 26 there are differences that are worth discussion.

But prior to the discussion, the significance differences of all of the good and average FSI scores taken diagonally (top-left to bottom-right) and calculated using the Wilcoxon Matched-Pairs Signed-Rank Test is high: a two-tailed probability of .0357. This means that the diagonal FSI differences between the good and the average students in the L2 tests are almost unlikely due to chance alone: the diagonal FSI scores of the good readers are consistently difference than the FSI scores of the average readers. The two-tailed or nondirectional hypothesis is used simply because the direction of the FSI scores differences could not be predicted. Thus, the result of the Wilcoxon test on the diagonal FSI scores was significant at the $p < 0.05$ level. The good readers consistently reflects more comprehension answering strategies on each corresponding QT and DT than the average readers and the differences are not due to chance alone.

Advancing the argument, regardless of the students' reading ability, inconsistencies are found in the FSI loadings between the two languages (see the diagonal FSI scores in Tables 26 and 27). The inconsistencies are not only discussed in the context of the diagonal FSI scores of the L2 tests but are also analytically compared with the equivalent FSI diagonal scores in the L1 tests as seen in Table 26. Table 28 (see the next page) summarises the FSI scores differences in the L1 and L2 tests for both the above average and average readers.

Table 28: A Comparison of the Diagonal FSI Scores Between the Above Average (Good) and Average Readers in the L1 and L2 Tests.

		<u>Language of Texts and Tests</u>	
		<u>L1 (Bahasa)</u>	<u>L2 (ESL)</u>
<u>Question Type</u>			
W	G	20	38*
	A	24	32*
WIC	G	42	29*
	A	20	25*
L	G	70	42*
	A	53	25*
ISS	G	65	60*
	A	55	26*
IMS	G	33	45*
	A	35	31*
M	G	26	32*
	A	06	14*
S	G	20	21*
	A	04	03*
J	G	76	73*
	A	80	80*
<p>Note: G: Good students</p> <p>A: Average students</p> <p>* P < .05 (Wilcoxon Matched-Pairs Signed Ranks Test)</p> <p>The diagonal FSI scores are taken from Tables 26 and 27 respectively. Each FSI score is the percentage of responses for that particular corresponding discourse type (DT) and question type (QT) only.</p>			

An Integrative Discussion Interpreting Tables 26, 27 and 28.

In the beginning, it was expected that the diagonal FSI scores would be higher in L1 than L2 tests because if the diagonal scores are low in L1 then it shows that the students are using the other sources of information to substantiate their reasoning processes. So one way of interpreting the L1 FSI scores is by assuming that a student can use information from the same DT itself because he/she has more knowledge in that particular DT skill. In responding to the L2 tests the student may need extra information sources, particularly the (*ISS*), (*IMS*) and (*WIC*) DTs because he/she may not be good in vocabulary. So, the expectation was that the diagonal scores in Table 28 would be lower in L2 than in L1. But that did not happen. What was found was that the diagonal FSI scores were higher in L2 than in L1. Interpreting this phenomenon is not easy.

Table 28 summarises the comparison between the diagonal FSI scores of the good and average readers in the L1 and L2 comprehension tests. It must be stressed that the number of available discourse units for both the L1 and L2 FSI scores in Table 28 is small. This is particularly true when interpreting the L1 FSI scores, which were derived from only 3599 discourse units. The L2 FSI scores were derived from 4243 discourse units. Without doubt this increases the risk of no significant difference in terms of the FSI scores between the good and the average readers in L1. Although in some cases the percentage difference in Table 28 is quite high, as seen in the Bahasa column, the differences are not significant because the actual number of discourse units is relatively low and thus the findings are not reliable enough to be statistically significant. By being not statistically significant means that the distribution of the FSI scores of the good and average students in L1 is widely spread than in L2. In other words, the distribution of the FSI scores between the good and the average readers in L2 are closer to the mean percentage of each of the FSI score and this makes the L2's FSI scores for the two groups of readers to be statistically significant. Nevertheless, the Wilcoxon Matched-Pairs Signed Ranks Tests were performed to find the reliability of the FSI diagonal scores for the good and average readers in both the L1 and L2 tests (see Table 28 for the significant difference of the two languages). But what is important is the patterns of the distribution of the

FSI scores in both the L1 and L2 tests, and these need further explanation.

Looking at the diagonal FSI scores for the (*W*) skill in L1 (see Table 28), the good readers are scoring slightly less than the average students. What was expected was that the good readers should have higher FSI scores in the (*W*) skill than the average readers. For the (*WIC*) skill in L1, 42% of the good readers' FSI scores are on the diagonal line compared with only 20% of the average readers. Initially, this phenomenon was predicted: that the average readers are 'moving away' from the diagonal because they use the other DT skills more than the good readers when responding to the (*WIC*) QT. This 'moving away' phenomenon is supported not only by the higher FSI scores of the good readers in the (*WIC*) DT but also in the (*L*), (*ISS*), (*M*) and (*S*) DTs. The point is that in the (*WIC*), (*L*), and (*ISS*) DTs the good students score higher in the FSI than the average readers.

In the L2 tests (see Table 28), the good readers' FSI scores in DT (*WIC*) are higher on the diagonal than the average readers. This was predicted. First of all, comparing the FSI scores for DT (*WIC*) in both languages, the scores for the good readers in L2 are much lower than in L1: 29% as opposed to 42%. Thus, as predicted, in L2 the good readers were 'away' from the diagonal and are assumed to interact with the other skills. This prediction is further supported by the L2 (*L*) FSI score of 42%. This is to say that when the good readers are working in the ESL context they need to go beyond the (*L*) skill in order to talk about QT (*L*). The average readers went even further off the diagonal when responding to the same QT (*L*). The crucial point is that the FSI scores of the average readers for the DTs (*L*), (*ISS*) and (*IMS*) in L2 are much lower than for the corresponding DTs (*L*), (*ISS*) and (*IMS*) in L1. The FSI score of the good readers for DT (*ISS*) in L2 is not strikingly different from the DT (*ISS*) score in L1.

Looking at the QT (*W*) of the L2 tests, the above average readers are reflecting only 6% more information than the average readers (see Table 28). But in terms of the concentration of the FSI scores for that particular QT across the two languages, there are sharp increases of concentration from L1 to L2 for both groups of readers. That is, by looking at Table 28, the increases are from (*W*) 20%

to (W) 38% for the above average readers and from (W) 24% to 32% for the average readers. This pattern of increase is not to say that there is a direct causal link between vocabulary and comprehension: increased reflection of the (W) skill means increase in comprehension. But what can be learned from previous research, which goes back to the early part of this century, in the context of the relationship between vocabulary knowledge and comprehension is the fact that knowing more words is statistically related to better comprehension of text (Ruddell, 1994).

In terms of attempts to establish the direct link between vocabulary knowledge and comprehension, Ruddell (1994, p. 414) claims that contemporary research efforts on this issue '... have been equivocal and inconclusive...(and) from the evidence available, the most we can say with assurance is that sometimes it does and sometimes it doesn't...'. In the context of the discussed evidence of the increase in the reasoning process of the (W) skill in the L2 tests in comparison with the L1 tests, this suggests that all the readers, regardless of the groups, verbalised more reasoning strategies for that particular skill in the L2 tests. Interestingly, the above average readers participated more actively in the (W) QT than the average readers. Still, in the context of this comprehension-process analysis, it is not known whether the increase in reflection observed in the vocabulary skill in the L2 tests as opposed to the L1 tests reflects a problem faced by the readers in understanding the meaning of the tested words. But what is clear is that the evidence from the diagonal FSI scores in both the L1 and L2 tests suggest an idea that contextual information facilitates the construction of meaning of the tested (W) skill. Still, differences in individuals' ability in using or not using contextual clues in answering the (W) QT are not tested. The evidence seems to suggest that there is a critical link between the (W) skill and the other remaining comprehension skills. This will be discussed later in the context of the Stanovich's (1980) interactive-compensatory model of reading.

It is interesting to note that when responding to the (W) QT in the L2 tests (see Table 27), the average readers show a sharp decline in using DT (WIC) from the L1 test scores (see Table 26): from 35% in L1 tests to 17% in L2. But, in comparison, for the above average readers or proficient readers, there is no sharp

decline observed. That is, for the same QT, the decline is minimal: from 36% to 30%. For the same (W) QT, the average readers consistently maintain their reliance on the (J) DT: 13% of the reasoning processes for the (W) QT in both the L1 and L2 tests are based on the (J) DT.

The above average readers are once again more reflective than the average readers in the (WIC) QT. But when looking into the (WIC) QT across the two languages, the above average readers are found to reflect considerably less in L2 than in L1. This is in sharp contrast with the average readers who show a slight increase in the FSI of the (WIC) questions in L2 from the L1 tests: from 20% to 25% (see Table 28). But, the average readers, when responding to the (WIC) QT, are found to be reflecting considerably more on DT (W) than the above average readers: 40% against 25% (see Table 27). This probably suggests that the average readers are very concerned with the meaning of the words or vocabulary or alternatively that the (WIC) QTs are causing them problems. This heavy reliance on the (W) DT for the (WIC) QT by the average readers is in sharp contrast with the above average readers who seem to show an increase in using other DTs in the L2 tests, apart from the (W) DT, such as DT (L) 7%, DT (ISS) 13% and DT (IMS) 9% (see Table 27). Comparing the different concentrations of the FSI scores of the (WIC) QT between the two groups of readers suggests that the good readers are seen to be more flexible than the average readers in terms of reflecting more on the context of the texts rather than heavily relying on vocabulary skills as reflected by the average readers.

In responding to the L2 (L) QT, both the groups shows a remarkable decline in using that particular skill from the L1 tests. As seen in Table 28, the above average readers' decline is from 70% to 42%, and the decline is from 53% to 25% for the average readers. In the case of the good readers the decline is compensated for by increased reliance on the other skills, notably on the (ISS) (25%) and (J) (18%) (see Table 27). The same increase in reliance on the other skills is also observed in the average readers: (ISS) 17% and (J) 30% (see Table 27).

In terms of the DT (*J*), both groups of readers reflect heavily on that skill alone, the scores being above 70% (see Tables 26 and 27). Still, the good readers rely slightly less than the average readers on this skill and this slight decrease is compensated for by an increase in reflection on the (*IMS*) skill: from 19% in L1 to 25% in L2. In the case of the average readers reflection on the DT (*J*) skill is consistently high in both languages but the patterns of the other DTs for that QT (*J*) are slightly disrupted: when responding to the DT (*J*) in the L2 tests, the average readers tend to use DT (*W*) 3% , which is not reflected on at all in the L1 tests.

In QT (*IMS*), the average readers talk more on DT (*J*) in the L2 tests than in the L1 tests. Almost 50% of the reflections for that QT is concentrated on DT (*J*) (see Table 27). Again, the average readers, as for the QT (*IMS*), reflect on the use of DT (*W*) 3%. This usage of DT (*W*) for QT (*IMS*) in the L2 tests is practised not only by the average readers but also by the good readers as well. In fact when comparing DT (*W*) and DT (*WIC*) in Tables 26 and 27, especially when one looks from QT (*L*) downwards, there is an increase in the usage of these vocabulary skills in the L2 tests in both groups of readers. This increase is particularly notable when both groups of readers respond to QT (*M*): the increase in the average readers is from 2% to 14% and for the good readers it is from 2% to 7%. Probably, the questions asked for the (*M*) skill in the L2 tests are causing problems to the students and they have to reflect on DT (*W*). As an example, the following discourse units taken from both groups of readers illustrate the problem of not understanding a particular word and the strategies used to compensate for the deficiencies.

Text: Si Pintar

Language of comprehension test: English

Subject: R. R. (Good reader)

Discourse units:259-272. (see Appendix D (viii)).

-
259. E: OK. Now number 17. What do you understand by the “shafts of gold” in paragraph 1 ?

260. S: The light by the sun.
261. E: The light, how is it? What kind of light is that?
262. S: So, it said a 'It lay in a pool of shadow...' (Paragraph 1).
263. E: Aha?
264. S: '...with only shafts of gold where the sun pushed through the trees.' (Paragraph 1). It means the sun that shown (shone?) a bit on the pool.
265. E: Shown (shone?) a bit on the pool?
266. S: On the pool of shadow.
267. E: Pool of shadow?
268. S: Slightly.
269. E: How is it? Is it really a aa "shafts of gold"?
270. S: 'Shaft', I am not so sure what is the meaning of 'shafts'. But I understand that it is the light shone by the sun.
271. E: The light shone by the sun?
272. S: Mm.
-

In the above discourse units, the good reader (see Discourse unit 270) mentioned his uncertainty of the word 'Shaft' and thus he is coded as reflecting on the DT (W). This deficiency is compensated for by relying on his inferences from DT (ISS) as indicated in discourse unit 264.

In another example, when responding to the same QT (M), the average reader failed to diversify his reasoning strategies in finding alternative action(s) in responding to the question. This is best illustrated by the following example:

Text: Si Pintar

Language of comprehension test: English

Subject: M. F. J. (Average reader)

Discourse units: 347-358. (see Appendix E (vi)).

- 347 E: OK. Number 17. 'What is meant in the story...' Do you know what is 'meant'?

348. S: Meaning (makna).
349. E: The meaning of 'shafts of gold' in paragraph 1, mmm? Have you seen that before? You couldn't imagine its meaning?
350. S: Ha.
351. E: Right? This 'shafts of gold'?
352. S:(No verbal response)
353. E: What is your guessing?
354. S: (No verbal response)
355. E: Don't know, right?
356. S: Ahm.
357. E: But of course you know what 'gold' is?
358. S: Hmm.
-

What can be inferred from the above discourses is the fact that the average reader failed to trigger or mention any alternative reasoning strategy in response to the demand of the QT (*M*). He seemed to be 'paralysed' within the hemisphere of DT (*W*) alone and failed to mention alternative responses.

Although the above two examples may not represent the real truth of the matter, the increase in reliance on the vocabulary skills in the L2 tests as compared to the L1 tests especially from QT (*L*) to QT (*J*) is worth discussing. It seems that automatic word recognition is crucial in text comprehension and the more the students could master the meaning of words, contextualised or decontextualised, the better they would become at comprehending the texts. It also means that once the meaning of the words are comprehended automatically then more time could be spent in inferencing the proper meaning of the question. But, even with a deficiency in L2 lexical meaning, as demonstrated by the good reader, there are alternative monitoring strategies in checking the solution. So, this study suggests that comprehension monitoring strategies may have more to do with reading comprehension proficiency than the language backgrounds of the readers. It seems that, despite the deficiency or uncertainty of the meaning of the word 'shaft', the good reader was able to verbalise his regulation of the strategic plans to find the best answer. The average reader was unable to suggest

any plan to solve the problem. As said earlier the average reader was paralysed with the question itself. It is not known whether the average reader was cognitively unaware of what to do or simply unwilling to reflect the truth of the matter. But what is clear is that the good reader was flexible and adaptable to the need of the comprehension question. This efficiency is not noted in the average reader.

In relation to the issue of word recognition, it is pertinent to reflect on the current debate of Goodman's (1967, 1968) view of reading as a psycholinguistic guessing game. Harrison's (1996, pp. 9-17) analyses of the criticisms made by many current reading scholars on the inadequacies of Goodman's guessing game are centered on four points and one of them, related to the context of this study, is on the statement that '*Good readers are not dependent on context for word recognition*'. Although it is sensible for Goodman to claim that in the quest for meaning good readers are efficient in the use of context, current views on word recognition emphasise the fact that good readers are better than poor readers at automatic, rapid and context-free word recognition. But when the good readers' automatic word recognition fails, they resort to the interactive-compensatory mechanisms as illustrated above by the good reader coded as R. R. In this study, the author feels that rapid, automatic word recognition in reading comprehension in L2 texts is crucial to fluent reading comprehension but at the same time automatic word recognition alone need some kind of trade-off if comprehension breaks down. This trade-off phenomenon, as exemplified time and time again in Tables 26 and 27, are very much dependent on the readers' abilities to contextualise their comprehension monitoring. The good and average readers need to increase their rapid L2 automatic word recognition and master compensatory strategies if comprehension breaks down.

In the case of the (S) skill, the good and the average readers' FSI scores in both languages are more or less the same. It is also observed that the FSI scores for this skill in both languages are less than 21%. This is not surprising because logically when responding to this type of skill, the students would rely heavily on making inferences and judgements. This heavy reliance on the (IMS) and (J) skills are noted in both languages. For both groups of students and in both

languages the FSI scores of the (*IMS*) for QT (*S*) fall between 44% and 54%. This high percentage reflects the dependence of both groups on the (*IMS*) skill when responding to the QT (*S*). Whether poor readers would behave in a like manner is a matter for future research.

Conclusions Drawn From Tables 26 and 27.

To summarise, in the light of the above discussions several important conclusions can be drawn from Tables 26 and 27. First, by referring to the third hypothetical construct, the evidence discussed seems to suggest that there are marked differences between the good and the average readers in terms of the distribution of the discourse unit (FSI) in both the L1 and L2 comprehension texts and tests. In other words, both in the Bahasa and in the English comprehension tests, the above average readers consistently show higher FSI scores in almost all specific QTs, shown diagonally in Tables 26 and 27, than the average readers. The gap in the FSI scores between the two groups of readers becomes more apparent in the L2 tests: the good readers consistently reflect higher FSI figures than the average readers except in the (*J*) skill. Second, in a similar manner to the L1 tests, the good readers' responses or DTs in L2 (as shown by higher FSI figures than the average readers) were very much focused on the QT. This may suggest that, for each QT, the good readers reflected less information from the 'neighbouring' DTs and this phenomenon could suggest that the good readers are able to gear their answers specifically to the need and context of the questions. This higher-focus-phenomenon does not mean that the average readers are more able to spread their reasoning capabilities to all the other DTs than the good readers. It suggests that the good readers have less need to do so.

Third, since the reasoning strategies of the above average readers are repeatedly very much more 'bonded' to each QT than those of the average readers, this 'preferred-and-most-often-used' phenomenon reflects a kind of efficient reasoning strategy. In contrast, the average readers showed a less

'bonded' approach and a greater variability in reliance on the other DTs particularly in the consistent use of the (*J*) DT. It is not known whether the less 'bonded' patterns of the average readers reflect a kind of less efficient answering strategy.

Finally, the diagonal results from Tables 26 and 27 seem to suggest that in both the L1 and L2 tests, the good and average readers use the same comprehension answering strategies, but the good readers are more consistently focused or bonded to each QT than the average readers.

Important Conclusions Drawn from Table 28.

Based on the discussion on Table 28, there are two main points which must be stressed. First of all, the FSI scores of the good readers in the L1 and L2 tests on DTs (*WIC*), (*L*), (*ISS*), (*IMS*) and (*M*) are concentrated on the diagonal more than those of the average readers. The second point is that the FSI scores of the good readers in L2 on DTs (*WIC*), (*L*) and (*ISS*) are lower than in L1. The fact is that in the L2 tests even the good readers are behaving more like average readers. This can be seen from the DTs (*WIC*) (29%), (*L*) (42%) and (*S*) (21%): in all these three skills the good readers go beyond the diagonal. What we are seeing from the diagonal FSI score patterns is an interactive-compensatory comprehension process. This is parallel with Stanovich's (1980) idea of the interactive-compensatory word recognition process.

According to Stanovich (1980) good readers are far better at automatic context-free word recognition than poor readers. Poor readers need to use larger contextual facilitation than do good readers in word recognition. In the case of reading comprehension answering strategies, a paralld pattern as in Stanovich's (1980) emerged from the diagonal FSI scores in Table 28: the good readers can get the information they need for comprehension from the 'local' or corresponding DT area, the average readers need to go more 'broadly' into DT (*J*) in making sense of the questions. This 'broad reliance' is shown by the high FSI scores in DT (*J*) (80%) in both languages. In both the L1 and L2 tests, the average readers used wider contextual facilitation for QT (*W*) than the good

readers. This is shown by the high FSI scores on DT (*J*): 13% in L1 and 13% in L2. This means that the processing capacity of the average readers that is free for comprehension is affected. In a different manner, when responding to the same DT (*W*), the good readers shown almost no reliance on DT (*J*): 0% in L1 and 3% in L2. So, the good readers have more processing capacity left for comprehension processes. It is interesting to add that in looking into the different diagonal FSI scores in L1 and L2 tests, in the L2 tests every reader is behaving more like an average reader: even the good readers need to go beyond the 'local' DT in the search for meaning (see DTs (*W*), (*WIC*) and (*L*) in the L2 diagonal scores).

Chapter 10

Conclusions and Recommendations from the Part I and Part II Studies: Some Implications to the Reading Comprehension Syllabus and Classroom Practice.

The two studies were designed specifically to seek a better understanding of the development of reading comprehension in Malaysian primary and secondary schools. In doing so, the first of the studies begins by addressing the classical issue of whether reading comprehension can be broken down into a number of distinct subskills. It then replicates and extends the work of Lunzer and Gardner (1979) in selected Malaysian primary and secondary schools.

The unique aspect of this small survey study is that the issue of hierarchical subskills in reading comprehension is not only replicated in and adapted to Bahasa Malaysia or as it is commonly called, Bahasa, the national and official language of instruction in Malaysian schools, but also includes English which is the official second language of instruction and communication. Such an extension of Lunzer and Gardner's (1979) study into English as a Second Language (ESL) is seen as necessary in exploring and advancing the issue into the paradigm of teaching and learning reading comprehension in Bahasa and ESL.

There is no doubt that in this study the designed comprehension tests only measured students' recognition of answers to each of the eight comprehension subskills. Factor analysing the written products, within the hypothetical sphere of the assumed hierarchy of subskills, allows the author to identify and infer the existence of a hierarchy of subskills in reading comprehension. As in the case of Lunzer and Gardner's (1979) findings on the same issue, the author too, could find no evidence to support the existence of discrete subskills in reading comprehension in Bahasa and English. This conclusive finding is consistent with Smith's and Goodman's approach: reading is a unitary process. With this unitary view, the author feels that there is a necessity of extending the traditional comprehension test into current reading research paradigm: what good and

average readers do in developing their comprehension answering strategies. With this 'process' view, the second part of the thesis attempts to explore and understand the reading comprehension answering strategies of good and average readers in selected lower secondary schools in Malaysia. The central concern of the process-oriented Part II study is to explore a new frontier in understanding comprehension answering strategies within the eight subskills paradigm. In a subtle manner it seeks to substantiate the product-oriented method of the Part I study with a process-oriented paradigm that could broaden the horizons of understanding reasoning processes in reading comprehension. Both the product and process oriented research approaches are useful in understanding reading comprehension if one wants to get a holistic view of the complexities of the reading process. In a strict sense Part II of this research seeks to explore three underlying hypothetical constructs through analysing 16 case studies.

It is understood that any attempt to develop applied reading research and theory must not ignore the importance of basic reading research in confirming and developing the contemporary reading theories. With this in mind, the second part explores the development of comprehension answering strategies by means of face-to-face interview. It is believed that the retrospective verbal responses could provide the author with ample opportunities to closely examine patterns of the students' practices and reasoning processes viewed within the stated eight categories of subskills. In a sense it inevitably goes well beyond the implementation of reading comprehension tests by attempting to infer and to code each discourse unit of the verbal data into one or more of the eight categories of subskills as used by Lunzer and Gardner (1979). Analysing the coded verbal responses or protocols allows us to gain insight into the paths, ways or patterns in which the selected group of students as a whole, or as categorised into two reading ability groups, organized the information gained either from the texts or from another source when responding to the comprehension tasks. Among other things, the main aim of the analysis of the 16 case studies is to understand the relationships of the discourse types and the question types in the comprehension tests, the patterns of their relationships between the answers of the good and the average readers and the different

effect(s) of the language of the comprehension tests on the distribution of the discourse types and the question types in the two groups of readers.

This basic research should be regarded as an attempt to explore the above-mentioned hypothetical basic differences and although it is suggested by Bernhardt (1991, p. 227) that '...second language literacy research should recognize itself as an entity distinct from first language literacy research' the researcher feels that this study could cast some mind-opening findings on the reasoning strategies of the readers in L1 and L2 through the retrospective interviews. In other words, this basic research is useful for exploring and improving our understanding of reading comprehension not only in L1 but also to a lesser extent for exploring and expanding L2 reading research into a principled and unique research entity.

The researcher suggests that from the findings of the Part II study there are some basic common reading comprehension answering strategies used by the good and average readers when responding in both the L1 and L2 comprehension tests that may suggest a unique reading research paradigm that transcend Bernhardt's suggestive distinct entity. In other words, the findings from Part II suggest that there are some common reading comprehension behaviour patterns in both the L1 and L2 reading comprehension tasks that are symbiotic in perspective rather than strictly unique to each language. As an example, in responding to all the comprehension tasks in both L1 and L2 texts and tests, the students displayed constant routines of 'hanging-on' to a few Discourse Types and regulating their reasoning processes in a 'compensatory-interactive' manner (see the discussions for Tables 23, 26 and 27). These routine characteristics are by themselves 'unique' in their own right. Thus, in teaching and assessing reading comprehension, reading teachers should have a symbiotic view when looking into the nature of reading comprehension strategies and should understand the capacities of readers to overcome language barriers in their attempts to make sense of the comprehension tasks.

It is important to stress in a thesis such as this that understanding the theories and complex processes involved in reading comprehension does have some practical implications in the design of the reading comprehension curriculum, teacher training and classroom practice. There is no doubt that this thesis, apart from testing the said hypothetical constructs, aims to draw attention to the issue that may have been taken for granted or seen as 'unproblematic' by the affected groups.

10.1 Evidence of Unitary Subskills in Reading Comprehension from the L1 and L2 Tests and its Implications for the Reading Comprehension Syllabus and Classroom Practice.

This thesis began with the classical debate on the issue of the holistic versus the subskills approach in the teaching of reading comprehension. It began with the simple question of whether reading comprehension is a unitary competence of subskills or one that can be broken down into separate subskills. Then it continues with the replication and extension of the hypothetical construct of Lunzer and Gardner (1979), that reading comprehension can be broken down into a number of distinct subskills, in the L1 and L2 tests. The discussed evidence from the L1 and L2 data analyses proved that the eight categories of reading comprehension subskills in both L1 (primary school students) and L2 (lower secondary school students) tests behave in a holistic manner. This 'global act' of reading comprehension in both languages is more than just the sum of its parts: the assumed separate and different subskills are in fact interrelated or integrated in most of the factorised data. Clearly, these findings support the findings and views of Lunzer and Gardner (1979), Goodman (1970), Smith (1971) and other proponents of the holistic camp.

The evidence gained from this study and other similar studies raises the question for all the affected groups of rethinking the nature of teaching reading comprehension. In the first place, there is no doubt that most of the reading comprehension teaching practices in Malaysia are focused on the drill-and-practice methodology. As an example, the findings from a survey conducted in 1982 by the Federal Inspectorate of Schools in the state of Selangor, Malaysia, on the teaching of reading comprehension by lower secondary school teachers showed a clear manifestation of the traditional methods of reading instruction. Among others, Mustapha (1994, p. 5) found that the majority of the English reading comprehension lessons observed from the survey were focused on improving:

- i. ... the pupils' pronunciation and vocabulary;
- ii. ... their reading-aloud ability;
- iii. ... their productive rather than their receptive skill.

Mustapha (1994) pointed out that these findings were also observed in the follow-up nation-wide survey on a similar topic. In the world of teaching reading comprehension the above findings are not shocking because the traditional drill-and-practice teaching model assumed an hierarchical order of subskills that must be mastered. The mastery of all the subskills is equated with reading comprehension (Dole et al. 1991). Tierney and Pearson (1994, p. 501), on the current observations on the trend in teaching reading comprehension, found that:

... there is and has been a lot of comprehension testing and practice (students working by themselves on worksheets or answering questions) and a great deal of informal assessment ... In most lessons, students are given passages to read. During or after reading the passages, teachers ask questions (either orally or via a worksheet). Any discussion of responses focuses on finding a right answer. In terms of skill acquisition, a high premium has been placed upon separate objectives unrelated to any comprehensive model of reading comprehension or learning and clustered around curriculum objectives or arbitrarily defined skill categories (e.g., literal, inferential, and evaluative comprehension) that give little attention to the role of a reader's background knowledge and the importance of improving a readers' abilities to learn how to learn.

The above traditional trends in teaching reading comprehension are also echoed by the national survey's findings in parallel situations. Mustapha, the chief Inspector of Malaysian Schools, (1994, p. 8) also noted that among other things, ' ... [from the 1983 national survey in 64 Malaysian urban and rural lower secondary schools the] reading ability and interest among students in secondary schools all over the country were generally low. In most EL [English language]

classrooms reading involved the mere decoding of text to obtain meaning. [From the 234 reading lessons observed] ... Little attention was paid to interacting with the information contained in the text.' Basically, what Mustapha (1994) observed is the lack of attention given by the reading teachers on fostering the students' comprehension monitoring abilities, also called metacomprehension abilities. This lack of focus on teaching comprehension monitoring by reading teachers was also observed by Schmitt and Baumann (1990): a total of ten primary school teachers (Grade 1 to Grade 6) were observed on the teaching of metacomprehension skills during guided reading of basal reading instructions. Schmitt and Baumann (1990) found that too little attention was given to teaching students on metacomprehension strategies. Activities that are known to foster comprehension monitoring as observed by Schmitt and Baumann (1990) are: previewing material; activating prior knowledge; determining text characteristics; determining purpose for reading; generating questions; predicting; verifying predictions; recognizing a comprehension breakdown; and employing repair strategies.

Dole et al. (1991, p. 246) also acknowledge the importance of comprehension monitoring in reading instruction and say that much research has proven that 'Good readers are better than poor readers, not only at reading but also at monitoring, controlling, and adapting their strategic processes while reading. Poor readers, by contrast, are much less aware of problems that exist and less able to solve problems even when they are aware of them.'

This lack of attention to interaction with the text, as observed by Mustapha (1994), implies that reading comprehension has been perceived and taught by the teachers as mere application of isolated subskills and the accumulated mastery of all the subskills by the readers has been thought to mean that the readers have mastered reading comprehension. If such an assertion exists in the primary and lower secondary schools in Malaysia then a shift in the reading comprehension paradigm from product to process oriented is much needed. Recent research developments in reading comprehension are no longer focusing on the product of isolated skills but on the dynamic of '... meaning acquisition process, necessitating awareness and control of very involved

reasoning process.' (Myers, 1991, p. 259). In other words, reading is thought of as active process and the readers are thought to make efficient use of whatever strategies they can in the attempt to make sense of the text or the printed information (Goodman, 1970; Smith 1973; Miller 1963). This does not mean that the current Malaysian reading comprehension syllabus for the lower secondary schools is of the old traditional regime. In fact it echoes the current theoretical understanding of reading research and in particular emphasises the concept of communicative approach in language skills. In the '*Compendium: A Concise Guide For Teachers of English* (Kementerian Pendidikan Malaysia, 1989a, p. 38) reading is defined as:

a 'receptive' skill, this means that when we read we receive (i.e. read and understand) what a writer has written. But although reading is a 'receptive' skill this does not mean we are passive when reading. We need to react to what we read in order to understand the meaning of what has been written.

What it means for the reading teachers is that clear understanding of the theories, models and processes of reading is necessary and this is particularly true in mastering the teaching of reading comprehension strategies either in L1 or in L2 classroom reading instruction. Again, this is not to blame reading teachers in English, as observed by Mustapha (1994, p. 7), for the fact that '... [from the survey] a large number of teachers failed to help students develop proper reading strategies and acquire comprehension skills necessary for processing information in reading texts. Instead too much emphasis was placed on getting the correct answers ... ' but to assert that reading teachers have to be properly exposed to current developments in the teaching of reading comprehension particularly in promoting active reaction to texts through metacomprehension activities.

Although the author has not been able to obtain equivalent information on the teaching of Bahasa Malaysia as is found in Mustapha (1994) on English, it does not mean that such low performances in reading comprehension in L2 in the lower secondary schools cannot be linked with reading performances in L1. In fact research evidence support the view that '... the interdependence between

second and first language reading skills is influenced by cognitive as well as socio-cultural factors, whose influence, moreover, may fluctuate over time ... [and] too little is known about the precise nature of these dynamic interactions.' (Hulstijn, 1991, p. 13). In a similar tone Wallace (1992, p. 22) says ' Reading is unitary process both because it cannot be adequately broken down into separate skills ... and because we draw on similar processing strategies in the reading of all languages,... '. What Hulstijn (1991) and Wallace (1992) suggest is the idea that regardless of the language of the text the cognitive processing strategies in making sense of the written words is interactive yet similar in behaviour. This interactive processing strategies across languages means that reading teachers need ' ... to maximize learners' potential by drawing on existing strengths in either the first or other languages.' (Wallace, 1992, p. 23).

Although this study is not directly investigating the issue of interdependence of skills between Bahasa and English as a Second Language (ESL) it could still be hypothesised that some sort of reading skills proficiency could be transferred from both languages. However, such interdependence of skills is important in understanding L1/L2 reading skills differences in both the skilled and less skilled readers. It may well be that poor performances in L2 comprehension are associated with poor performances in L1 comprehension but caution must be taken if one wants to apply any result from the L1 research to L2 reading instruction. Block (1992), in an attempt to explore the comprehension monitoring processes of 25 L1 and L2 college freshmen readers through think-aloud protocols, found that in terms of the differences in the comprehension monitoring of the readers, the reading proficiency of the readers is more crucial than the language backgrounds of the readers. Block (1992) found that the proficient L1 readers were more frequent and explicit in identifying the problems encountered in reading and in explaining their strategic plans than the proficient L2 readers. The comprehension monitoring processes of the less proficient L1 and L2 readers were not as good as the proficient readers of either language background.

Although the findings by Block (1992) concerned the freshmen performances from an urban college, the implications of her research seem to be

fairly straightforward. The Malaysian reading curriculum designer and secondary reading teachers in particular must take early action in planning and teaching the metacomprehension skills (comprehension monitoring processes) and avoiding the several shortcomings in teaching reading comprehension mentioned by Mustapha (1994). In fact the author observes, after several years of experience as a language teacher in teaching English for Academic Purpose (EAP) and English for Specific Purpose (ESP) to the undergraduate freshmen in one of the higher learning institutions in Malaysia and in teaching and observing the teacher-trainees-in-classroom-practice for the Postgraduate Diploma in Education (PGDE) for another higher learning institution in Malaysia, that there is a lack of focus in attitudes to teaching critical thinking in reading comprehension, and this aspect of learning should be fostered early in the secondary schools. Once they are taught adequate comprehension monitoring skills, such as DARTs (See Lunzer et al., 1984), students are likely to be more active critical readers in interpreting the content of the text to suit their own needs and will be better prepared for their learning in higher institutions.

The author's personal past experience as a campus residence-fellow and through everyday social interaction with the said candidates suggests that there is an 'aura of fear' or dislike of learning L2 which may be due to a lack of motivation, awareness or other unknown factors. Such problems in the candidates' perceptions of L2 learning should be tackled early or else they will jeopardise the achievement of successful qualifications. Although this personal experience merits further exploration and may not be a true reflection of all of the candidates, it reflects the need to look back into factors that could motivate and encourage readers in primary and secondary schools towards an early mastery of the metacomprehension skills especially in L2 reading. Hypothetically, low motivation in learning L2 may be linked to the ideological perceptions or inclinations of the language teachers teaching the language skills. Obviously, passive reading instruction should not be encouraged as a dominant approach. In fact in reading instruction in ESL, Block (1992, p. 337) suggests that '... readers should be made aware that questioning and monitoring is part of good reading, not the result of imperfect knowledge of their second language.' The current communicative-integrative approach in teaching Bahasa and ESL in the

secondary schools as recommended by the Malaysian Ministry of Education is very much in line with the active metacomprehension reading skills. The current Integrated Curriculum for the Secondary Schools (ICSS) in Malaysia, particularly in teaching reading skills in both languages, hopes to produce creative, skilful, critical and analytical readers (Haji Husin 1992). But if there is a persistent continuation of the product-oriented approach in the teaching of reading as observed by Mustapha (1994) then the desire to produce such critical readers will be problematic. Again, it can be hypothesised that if the L2 (ESL) teaching and learning reading methods become process oriented then the aura of fear could be reduced or eliminated.

The holistic nature of the assumed hierarchic 'subskills' in reading comprehension in the L1 and L2 comprehension tests highlights some practical implications for the design of the reading comprehension syllabus and classroom instruction. At the outset of the findings it must be noted that the comprehension texts used for the testing were narrowed to the narrative and biographical types and not the ones encountered in everyday life such as informative scientific texts. It may well be that future research expands the hypothesis into such texts. As far as this study is concerned, considering the robust findings of the 'global act' in reading comprehension subskills within the parameters of the tested texts, the assumed hierarchical subskills must be integrated and the reading comprehension curriculum should not be what is termed by Dole et al. (1991, p. 240) as 'an assembly-line model of skill acquisition'. This 'assembly-line model' can be traced back to the behavioural and task-analytic reading activities that existed half a century ago in the United States (Dole et al. 1991). Harrison (1994) noted that the latest reading curricula for England and Wales are also, to some degree, linked to the 'subskills' age-level performance. In Malaysia, as noted before, the assumed hierarchy of reading skills which are taken to be 'unproblematic' (Kementerian Pendidikan Malaysia, 1982) may lead to the 'assembly-line model' if they are not confronted. More seriously, the 'subskills' paradigm is noted by Harrison (1994, p. 84) to create problems in a situation when readers are observed:

[to] behave like experts when they are handling familiar text, and like novices when they are handling unfamiliar text. Readers perform well on a 'subskill' when they understand a passage, but poorly when they do not. So ability in reading is not constant: it is context-related.

Text familiarity and its effects on students' comprehension as stated by Harrison, especially in the context of L2 reading, was also proven by Safiah (1985) (See chapter 1.3).

In the context of Part I of this thesis, it is the intention of the author to bring out the distinctions between the passive traditional *skills* and the dynamic of *strategies* in teaching reading comprehension. Dole et al.'s (1991) review of related studies of the two terms is worth mentioning. *Strategies* in reading focus more on ' [first] intentional and deliberate plans under the control of the reader, ... [second] reasoning; readers use reasoning and critical thinking abilities as they construct and reconstruct evolving meanings from the text [third] flexibility ... readers modify strategies to fit different kinds of texts and different purposes ... [and finally] metacognitive awareness ... [in]...regulating and repair [of readers' understanding].' Whereas, in contrast to each of the four points above, traditional *skills* are observed to be based on ' ... [first] automated routines, ... [second, they] tend to be associated with low levels of thinking and learning ... [third] in reading pedagogy, [skills] connote consistency, if not rigidity, in application across a variety of texts ... [and finally] ... it is assumed that with repeated practice and drill readers would automatically apply the skills they learn to whatever they read.' (Dole et al. 1991, p. 242).

Active in-depth processing of the text, as suggested by Dole et al. (1991), was also echoed in DARTs (Directed Activities Related to Text) (see Lunzer *et al.* 1984). In general the activities in the DARTs are broadly categorised into two parts: Reconstruction and Analysis. What can be inferred from the DARTs activities is the fact that readers are encouraged to be more aware of the importance of using reading strategies and that teachers have to be properly organised, for example in terms of classroom management as recommended by

DARTs, in the appropriateness of the chosen passages to the students' prior knowledge and in working towards a proper understanding of the passages (Lunzer *et al.* 1984). Since the eight subskills are proven to be unitary in behaviour then how they behave in a 'global' manner is too important to be dismissed.

The primary and secondary language syllabuses (Kementerian Pendidikan Malaysia 1982, 1988, 1989b, 1989c, 1990 and 1995) stress the ambition of achieving a totally self-reliant reader as the ultimate aim in reading. Reading strategies, such as DARTs (which is 'global' in skills), have great significance in helping reading teachers to initiate, facilitate and motivate good, average and poor readers in managing and improving their own reading strategies and later become independent readers.

There are five reading strategies which tend to be used by skilled readers, as suggested by Dole *et al.* (1991, pp. 243-249), which have been empirically proven and can be effectively taught in the classroom:

- a) Determining importance (finding the gist, theme, topic or main point)
- b) Summarizing information
- c) Drawing inferences
- d) Generating questions
- e) Monitoring comprehension

These five reading strategies are more credible than the subskills because they are 'observable active processes' that focus on the regulation of the strategies. The traditional subskills processes are unobserved-automated-internal-processes that do not give a rich or informative account of the understanding of the texts read by the readers. If reading is viewed as an interactive and constructive process between the reader and the information in the text then there is a need to shift from the view of reading as '... a subjective and personal activity ... [to] a postmodern perspective, in which we place less value on "comprehension", and more on the processes of research and enquiry, on the reader's willingness and ability to explore texts, to compare version of reality, and to problematise in what

is read the concepts of "facts" and "truth" (Harrison, 1994, p. 88). DARTs activities, viewed by Harrison (1994) as a manifestation of the postmodern view of reading and which was first thought of as an extension of the Skinnerian behaviouristic school of thought, not only provide a radical and superior alternative view to the subskills-product-centered comprehension exercises but also encourage readers to be 'active-meaning-makers-at-work'.

In a sense, it is appropriate and reasonable to suggest that in designing reading programmes and reading activities for the teachers' training programmes, the strategic view of reading comprehension is enormously important. It is suggested that teaching the reading strategies and assessing the students' monitoring of comprehension strategies are bound to be difficult and one way of assessing the readers' strategies, as suggested by Harrison (1994, p. 89), is by considering a portfolio assessment system that could provide accountable evidence '... of reading activity, reading interviews, reading conversations and reading strategies ...' by the readers. This process or *how* evidence in the assessment is crucial in understanding and monitoring reading comprehension development at all levels of reading ability.

In line with the strategic view of reading comprehension is the fact that reflective reading programmes such as Directed Activities Related to Texts (DARTs) (See Lunzer et al. 1984), which focus on strategic critical analyses of texts, are more appropriate than the traditional 'subskills-product' tests of comprehension. Davies (1995), who was involved in the DARTs project, acknowledged that DARTs activities are appropriate for the learning of reading comprehension in all subjects taught in L1 and L2 at infant, primary, secondary and tertiary levels. In a sense, DARTs activities put forward a reflective concept of reading comprehension that is learner-centered. In the quest for meaning, the learners are systemised and demanded to be active decision makers and interrogators of information (see the two basic types of DARTs, Text Analysis and Text Reconstruction, in Lunzer et al. 1984). If integration of subskills is all important in students' mastering of reading comprehension, then those who make syllabuses and classroom practice must be aware of the importance of comprehension monitoring of texts as recommended by the DARTs project.

An attempt by the Malaysian Curriculum Development Centre (CDC) in promoting a process-oriented approach is worth discussing. Due to the national survey reports on teaching reading comprehension in ESL, the centre embarked on a preliminary project called the Interactive Reading Project in eight secondary schools in the state of Selangor. The 1993 project reported positive teaching and learning outcomes in many aspects such as teaching techniques, awareness of teachers of the reading skills (see Appendix B in Volume II), positive development by students in critical reading ability and more importantly the project improved the students' interest in reading (Mustapha, 1994).

To summarize, several important conclusions can be drawn in the light of the above discussion. First, unproblematic as it may be, teaching reading comprehension should not be viewed as the 'assembly-line' or hierarchical model that focuses on teaching individual subskills and later assumes that the mastery of each subskill will make a reader apply all the mastered subskills in any reading situation. This is far from true. In fact, it can be learned that the dynamic of reading comprehension subskills should not be viewed as an hierarchy of subskills and it is not true that without the mastery of the 'lower-order skills' a reader may not be able to respond to 'higher-order skills'. The comprehension subskills should be seen in the perspective of a 'global act' or in an interactive unitary dimension. This implies that reading teachers should not fix the assumed sequence of comprehension subskills. They should realise that the assumed 'higher-order skills' (e.g. *(M)*, *(IMS)* and *(J)*) could be taught, from the very beginning, simultaneously with the assumed 'lower-order skills' (e.g. *(W)*, *(WIC)* and *(L)*).

This 'global act' of teaching the subskills implies that each reader is unique in their understanding and teachers should interact with the readers and monitor their understanding of the texts. Second, in relation to the first suggestion, teaching reading comprehension should not be viewed as an individual and personal activity that focuses on the traditional style of teaching with emphasis on finding correct answers, pronunciation and oral reading. It should be viewed as reflective reading that encourages process-oriented instruction that can foster

the students' abilities to actively react to text and later become independent readers.

Third, it is important that more emphasis be given to teaching comprehension monitoring strategies that could improve thinking and reflection. In doing so, the teaching of the eight unitary subskills coupled with the monitoring strategies is useful in developing critical awareness of information in the text. Finally, passive reading comprehension instruction should be replaced by teaching active reading monitoring strategies and this is particularly true in the context of second language learning albeit that comprehension strategy research in L2 is still in its infant stage and should be continued and be the main theme of L2 reading research for the 1990s (Grabe, 1991).

The second part of this study attempts to look into comprehension monitoring strategies in L1 and L2 and expand our understanding of the reasoning processes involved in responding to the comprehension tests.

10.2 Findings and Implications of the Second Part of the Study for the Reading Comprehension Syllabus and Classroom Practice.

The second part of this study seeks to explore the reading comprehension answering strategies used by the L1 and L2 lower secondary readers within the stated definitions of the eight categories of subskills. It begins with three hypothetical constructs: first, there is a relationship in the distribution of the discourse units between the eight question types and the eight discourse types, second, there is a difference between the good and average readers in terms of the patterns of the discourse units, and finally, there is a difference between the good and average readers in terms of the distribution of the discourse units related to the language of the comprehension test passages and the language of the questions. The following discussions are centered on each of the three hypotheses.

Hypothesis One

The lengthy discussions on Table 23 in Chapter 9 aim to establish the fact that there is a strong relationship in the distribution of the discourse units between the QT and the DT. This relationship is strongly suggested by the high FSI scores of *Literal Comprehension (L)*, *Drawing Inferences from a Single String (ISS)*, *Drawing Inferences from Multiple Strings (IMS)* and *Forming Judgements (J)*. These high FSI scores can be seen diagonally in Table 23. They suggest that when the students are reflecting on their comprehension answering strategies they tend to talk more about those four skills than the (W), (WIC), (M) and (S) skills. It can be concluded that all the readers repeatedly utilise the four dominant skills, (L), (ISS), (IMS) and (J), in their reasoning capabilities and rely less on the remaining four skills. This does not suggest that the remaining four skills are useless. It is just that they are not reflected as commonly as the other four skills. What matters is the fact that all the eight discourse types are very useful in categorising the discourse units verbalised by the students.

The strong relationship of the DT and the QT is further supported by the interactive nature of the eight DTs in all the eight QTs. This is shown by the uneven FSI scores which are mostly spread thinly in all the eight DTs. Since the relationships between the QT and the DT are well established, secondary school reading teachers need to be aware of several points. First, in the assessment of reading comprehension, the teachers must be aware that students may rely heavily on several comprehension answering strategies as indicated by the four commonly-used DTs: (*L*), (*ISS*), (*IMS*) and (*J*). Whether the overwhelming reliance on these four categories of DT are appropriate in justifying their chosen answers needs to be fully understood especially in checking whether the answers are proper, rational and right. Second, when responding to any comprehension question type, students do behave in an interactive manner as far as the 'subskills' are concerned. This interactive manner implies that in reading comprehension answering strategies, the affected reading syllabus designer and reading teachers need to be aware of the importance of mastering the internal representation of the content of the text by the students. Such mastery will enable the readers to be flexible and critical in reasoning out their choice of answers. Reading activities such as the DARTs, discussed earlier, are useful in promoting critical readers and in mastering the internal representation of the texts.

The interactive behaviour of the various skills in responding to any QT also suggests that readers do need to be flexible and adaptable to the needs of the questions. This is to say that reading teachers must be aware of the importance of mastering the context of the text coupled with the personal experience of the students in regulating reading comprehension. The spread of the FSI scores indicates that none of the readers are 'stuck' or rigid in any hierarchic manner when responding to the comprehension questions. Indirectly this fact falsifies the 'cumulative subskills' paradigm. What is important is the fact that qualitatively poor readers are not as aware as good readers of the process of regulating, adapting and monitoring their comprehension (Dole et. al, 1991) and reading teachers should assist the poor readers to be flexible strategists. The four dominant DTs are the strategies used by the good and average readers in monitoring their comprehension tasks.

The first hypothesis is also extended in terms of the distribution of the FSI scores in the two languages of the tests. Again, in the context of the hypothesis, the FSI scores in both the L1 and L2 tests (See Table 24), seen diagonally and unevenly spread in an interactive manner, suggest that the relationship between DT and QT is well-founded. If one argues that deficiency in word recognition in L2 is a block to the understanding of the text then Table 24 suggests otherwise. When responding to QT (W) all the readers show a flexible interactive approach. This flexible approach is also demonstrated in all the other QTs. So, regardless of the language of the texts and tests, all the students seem to be making use of all the DTs, with the exception of DT (M) and DT (S), in making sense of their chosen answers. The diagonal FSI scores from Table 24 also suggest that there is a regular pattern of comprehension strategies that operates similarly for the L1 and L2 comprehension tasks. In simple terms, L1 and L2 reading teachers should understand that simply focusing on the products of reading comprehension, especially in L2 reading instruction, and ignoring the importance of building comprehension strategies that can create awareness in students of their metacognitive processes, may cause a setback in the process of creating independent readers. What is crucial is developing their process-oriented comprehension strategies, for example in the context of the eight DTs.

Hypothesis Two

Again as in the first hypothesis, the analysed data from Table 25 supports the alternative hypothesis: there is a difference between the good and the average readers in terms of the patterns of the discourse units. The reasoning processes of the above average readers are much more focused on each QT than those of the average readers. This shows that the good readers were able to activate their reasoning processes within the sphere of each of the QTs individually, relying less on other skills. This phenomenon strongly suggests that instruction in improving poor readers' comprehension monitoring strategies should teach this unique strategy of the good readers. In other words, reading teachers should improve the average and poor readers' awareness and knowledge in comprehension strategies because logically, an increase in awareness of strategic knowledge may improve pupils' performances on reading

comprehension tasks. Furthermore Paris, Lipson and Wixson (1994, p. 795) found that much research '... reveals that poor readers do not skim, scan, reread, integrate information, plan ahead, take notes, make inferences, and so forth as often as more skilled readers (do) ...'. Grabe (1991, pp. 392-393) also noted that there is evidence that in both L1 and L2 the '... young and less proficient students use fewer [comprehension strategies] and use them less effectively in their reading comprehension [than proficient students] ...'.

Another interesting point is that the average readers constantly rely more heavily on DT (*J*) than the good readers (See column DT 7 in Table 25). This surplus activation of the (*J*) skill suggests that the average readers had to assess the contents of the passages and judge them against their previous knowledge more often than the good readers. Whether this behaviour is time-consuming or effective and appropriate for each of the QTs is not known but it would be an interesting area for future research to study this over-use of DT (*J*) by average readers.

Hypothesis Three

The findings from Tables 26, 27 and 28 support the null hypothesis: there is a difference between the good and the average readers in terms of the distribution of the discourse units related to the language of the comprehension test passages and the language of the questions. Again, as seen in Table 28, although the two groups of readers used the same comprehension answering strategies, the above average readers were found to be more consistently focused on most of the QTs in both the L1 and L2 tests than the average readers. The fact that the good readers verbalised their comprehension answering strategies more frequently on each QT than the average readers reflects the importance of cognitive contextual awareness in mastering reading comprehension. Less proficient L1 and L2 students need adequate assistance from the teachers in helping them to evaluate, regulate and compensate their answering strategies in reading comprehension tasks. In fact, Baker and Brown (1984) suggest that effective comprehension monitoring instruction is necessary because its main aim:

... is to make the reader aware of the active nature of reading and the importance of employing problem-solving, trouble-shooting routines to enhance understanding. If the reader can be made aware of (a) basic strategies for reading and remembering, (b) simple rules of text construction, (c) differing demands of a variety of tests to which his knowledge may be put, and (d) the importance of attempting to use any background knowledge he may have, he cannot help but become a more effective reader. Such self-awareness is a prerequisite for self-regulation, the ability to monitor and check one's own cognitive activities while reading.

Dole et al. (1991, 243-249) suggest that there are five reading comprehension strategies that can be effectively taught in producing skilled readers: determining importance, summarising information, drawing inferences, generating questions and monitoring comprehension. Reading teachers should look into these five comprehension enhancement strategies that could be developed and later used or adapted by the readers when reading any kind of text.

In conclusion, there are several important findings from the analysis of the FSI scores in Table 28. First, Stanovich's (1980) interactive-compensatory word recognition model could be extended to another paradigm: there are interactive-compensatory comprehension process patterns, as proven by the behaviour of the eight 'subskills' which are complex yet interactive as seen in both the L1 and L2 diagonal FSI scores. Second, the diagonal FSI scores of the good readers in both the L1 and L2 tests on DTs (*WIC*), (*L*), (*ISS*), (*IMS*), (*M*) and (*S*) are higher than those of the average readers. This shows that the good readers are using more localised DTs for those skills than the average readers do. Third, the diagonal FSI scores of the good readers in the L2 tests on DTs (*WIC*), (*L*) and (*ISS*) are lower than in L1; and finally, in the L2 tests the diagonal FSI scores of the good readers on DTs (*WIC*), (*L*) and (*S*) behave more like those of the average readers: in these three skills the FSI scores of the good readers go

beyond the local DTs. The following discussion continues at more depth by looking into the quality of the comprehension answering strategies of the good and the average readers.

11.3 The Importance of the Acceptability of Students' Comprehension Reasoning Processes and Its Implications for Future Research.

Another implication of the Part II of this research is related to the issue of the verbalised critical reasoning of the students in response to the questions of the tests. To begin with, it is important to note the difficulties of comparing the verbal responses of individuals in the same ability group and across levels of competence. A respondent who verbalised more information on the *why* and *how* questions posed by the researcher yielded more '*categories*' than a respondent who talked less. This is unavoidable. Giving more responses could also mean being more alert and critical in reasoning out the written answers. But reflecting more categories for each question type is not necessarily synonymous with critical and acceptable answers. On the other hand, reflecting less categories for each question type does not necessarily mean that the answer is less critical, inadequate or unjustified. The issue that can be raised from the evidence gained from Part II of this study is the danger of regarding the quantitative differences of the FSI scores as the only criteria in judging the students' reasoning capabilities. What is more crucial is the issue of the acceptability of the reasoning process.

The evidence from Part II tells us that, in terms of reading ability the good readers are reflecting more thought (see the diagonal figures in Tables 25, 26, and 27) than the average readers. This situation in their reflection of answering strategies is almost consistent in all skills and in both languages. In any case, the respondents were allowed ample time to recall as much information as possible for every single answer. There are many possibilities that could have affected willingness and ability to reflect as much as possible on each given answer. The time of the day, anxiety levels, shyness, motivation and reflective abilities are some of the variables that were anticipated by the researcher. All the respondents were chosen for their openness and did not exhibit the shy behaviour that would jeopardise the intention of the interview. In other words only the talkative respondents were chosen from both groups. The following discourse units are samples of responses taken from each group. The patterns of the coded numbers could be identified and interpreted according to the context of the questions

asked. The discourses were on the English Si Pintar test. Two students, coded as R. R. (a good reader) and M. F. J. (an average reader) produced the following responses:

Questions 20 and 21 on '*Finding Salient or Main Ideas (M)* skill'.

R.R. : 2 1 2 4 7 1 7 2 8 5 5 8 5 5 8 5 3 4 4 3 5 5 5 5 5 5 5 5

M.F.J: 2 2 2 4 7 8 8 5 5 5 5 5 8 8 1 8 5 8 5 8 5 8 4 4 5 5 5 5 5 1 1 1 1

The first impression is that the categorised responses produced by M. F. J. are slightly more in number than those of R. R. Both readers equally recalled using skill (*IMS*), categorised as '5', more often than the other skills. M. F. J. recalled more on skills (*J*) and (*W*), categorised as '8' and '1' respectively, than R. R. The emerging issue here is whether producing more in each category reflects a critically acceptable reasoning process on the part of the reader in responding to each question type. This merits further discussion. The answers given to question 20 by the two readers will be explained in the context of the relevant discourse units.

Text: Si Pintar (See Appendix C (ii) in Volume II)

Language: L2

Question 20: Five words in the following list are important ones in the story because they tell us what the story is all about. Find them and underline them.

moon, pool, mule, cliff, rocks, meadow, wildcat, cave, forest,

grass, fight, ferns.

Note: The underlined words are the correct answers.

The followings are answers given to question 20.

Students' answers to question 20:

R.R: pool, mule, wildcat, cave and fight. (All correct as to the marking scheme)

M.F.J: mule, wildcat, cave, forest and fight. (Almost all correct)

Judging from the above list of categories (each number signified one complete verbal protocol), we can see that there is no marked difference between the two respondents in terms of reflecting their answers except that M. F. J. produced more verbal protocols than R. R. in the skills categorised as '5', '8' and '2'. Categorising the answers within the eight subskills was not easy. It involved careful reflection within the hemisphere of the text and the context of the questions and the jargon used by the students. The main aim of the coding task was to infer the intended meanings conveyed by the readers. Then, the reliability of the coding had to be checked.

Now, what can we learn from the categorised responses given by the two students? Both the students were judged to use almost all the subskills except subskill number 6 (*Interpretation of metaphor (M)*). Are the two students differ in terms of giving more information in each subskill? Both students in their reflections used subskill number 5 significantly, equally, but not subskill number 8. R.R. used significantly less of number 8 than M.F.J. In reading the Si Pintar text, it can be judged that simply by failing to acknowledge one of the best answers, in this case 'pool' as seen in M.F.J.'s answers, he does reflect a kind of 'unawareness' in his critical thinking as to the requirements of the question itself, but not in his understanding of the text as a whole. By choosing 'forest' instead of 'pool' as the answer he also raises another issue; whether the given answers, as a product, should be considered as totally wrong or partly right. But what is more important is the appropriateness of the critical reasoning process or the insights involved that have led the student to choose his answers. Below is a section of the reader's protocol for that particular question type. In this case the concentration of the discussion is on question 20.

M. F. J: (See Appendix E (vi) in Volume II)

453. E: Hm. OK. Number 20. Why these five are most important? Why 'wildcat' is important?
454. S: Wildcat' because aaa it make this story more interesting, yes, more interesting.
455. E: More interesting?
456. S: Haa.
457. E: Cave'?
458. S: The 'cave' is the place aaa a place aaa for Si Pintar to sleep.
459. E: A place for Si Pintar to sleep, right?
460. S: Ha.
461. E: Forest'?
462. S: This 'forest' is where this 'cave' is located.
463. E: OK. 'Fight'?
464. S: This 'fight' aaa the fighting between Si Pintar and this wildcat.
465. E: 'Mule'?
466. S: This 'mule' is Si Pintar.
467. E: The 'mule' is Si Pintar, right? The rest are not important? Moon' is not important?
468. S: 'Moon' is not important.
469. E: 'Pool'?
470. S: It's the same.
471. E: 'Cliff'?
472. S: (No verbal response)
473. E: 'Cliff', what is 'cliff'?
474. S: Hm,hm. (Meaning don't know)
475. E: Don't know? OK. Never mind, never mind. 'Rocks'?
476. S: (Whispering something as if don't know what it means)
477. E: 'Meadow'? 'Grass'? Not important?
478. S: (No verbal response)
479. E: 'Ferns' too?
480. S: No. (Not important)
481. E: Aaa, why the rest are not important?

482. S: Because all these are not helping him at that time and all that, right?
483. E: Not helping him?
484. S: Not really that, not really..even without all these in the story, the story could still be made.
485. E: Haa. Even without all these things, it could still be a story, right?
486. S: Ha.

Now, the argument for choosing 'wildcat' as the answer is not reflected in terms of who are the main actors. It is only based on the idea that the wildcat makes the story 'more interesting'. The statement is valid in its own right but is not a reliable and appropriate critical answer. 'Pool' is totally ignored as being not important and not much information is given as to why it was not chosen. So, it is as if there was a preconceived reason for not choosing 'pool', which should be chosen on the basis that it is the place of action which could mean either death or life for Si Pintar. He judges it as unimportant as far as the given choices are concerned. What can be inferred is that the ability to isolate the key points has not been fully mastered. This sort of ability requires the reader to probe and properly understand the chosen answers in the context of the story. The reasoning process in finding the answers should be interactive in nature: awareness of the need of the question and an active search for the best answers. Such interactive behaviour requires adequate understanding of the whole text and 'fine-tuning' of the judgement or decision made within the context of the story and the requirements of the question. So, this fine-tuning phenomenon could be one of the main factors in trying to understand the comprehension answering strategies of the good readers. Thus, a response with 'less' categories observed for a particular question type but which consists of 'fine-tuned' reasoning strategies is perhaps better than 'more' categories which are not appropriate or finely tuned to the interactive demands of the question and the context of the story.

This fine-tuning phenomenon also merits another dimension of discussion. A student can get a right answer but the insight or the quality of critical reasoning employed by a student in reaching an answer to a comprehension question, such as an (S) skill question, may reflect good, average or poor critical contextual reasoning. A suggestive term for this fine-tuning or 'appropriateness of answers' phenomenon is called Critical Reasoning in Reaching Conclusions from the Context of the Text, shortly, CRITEXT. This aspect of reasoning within the hemisphere of the eight reading skills is crucial in developing a proper understanding of the text.

In M. F. J.'s case, a teacher may be 'fooled' by the quantitative aspect of his answers but not by the qualitative especially when the teacher scrutinises the evidence of the student's responses using the context of the text, or CRITEXT. What is clear is the fact that the critical reasoning of the answers given by the student needs to be guided adequately so that more acceptable or appropriate reasoning strategies can be mastered.

Below is another verbal section on the same question taken from the other reader:

R. R. (See Appendix D (viii) in Volume II)

313. E: It doesn't sail in the sky. Next page. What make you choose the important words (in question 20) that tell us the story?
314. S: At first aaa first I must choose 'mule', the meaning of the 'mule', first. Explaining what the 'mule' (is), what does it mean.
315. E: Aha?
316. S: And then about the " cave ", the secret cave.
317. E: Aha?
318. S: After that about the 'wildcat'. Because these are the two important role in the what..
319. E: Two important roles; the mule and the wildcat.
320. S: Yeah. And then the wildcat wanted to attack the mule so there is there be a fight there and then..

321. E: Aha?
322. S: At last it ended in the pool it means maybe the wildcat died there.
323. E: Aha. So that is your concept, right?
324. S: Yeah.
325. E: Actors; 'mule', 'wildcat', 'fight', and then a place in the 'cave' and then rolled into the 'pool'. So what is wrong with 'moon', 'cliff', 'rocks', 'meadow', 'forest', 'grass' and 'ferns'?
326. S: These are like actually aa to explain the place, the rest are not important..
327. E: Amm?
328. S: Because it just want..like " cave " is more important to show where the struggle happened.
329. E: Aha?
330. S: Like 'rock' 'cliff' to show the scenery around there only.
331. E: So the rest are the scenarios?
332. S: Yeah.

From the above discourse on question 20 alone, it can be inferred that R.R. displayed a better CRITEXT in choosing 'mule' and 'wildcat' than M.F.J.. 'Mule' and 'wildcat' are considered as the two important actors in the story by R.R. because R. R. is able to isolate and differentiate the key points from the supporting points. With this 'finding the main idea' strategy in mind coupled with the mastery of the context of the idea(s) from the text, R. R. was able to critically choose the most important answers from the given choices. In a sense R. R. was able to fine-tune his reasoning strategies appropriately.

The above comparison and many others reveal the qualitative differences between the two readers in the application of contextualised critical reasoning in finding the best answers for the (S) questions. In fact the qualitative differences are also observed in the other skills. The above comparison illustrates to us the reasoning strategies that were taken into consideration by the students in responding to the question. The depth of reasoning used by the students is very

important in understanding how the good and the average readers select, analyse and reach decisions based on the needs of the question and the information in the text. The qualitative differences of the discussed CRITEXT performed by the two students are useful not only in making reading teachers understand the need for quality reasoning in reading comprehension answering strategies but also in assisting the good and the average readers to develop and sharpen their critical analysis of textual information. Thus, there is a need for reading teachers to interact actively with the students in demonstrating a few examples of CRITEXT. The purpose of the active demonstration of CRITEXT is to encourage meaningful discussions on the comprehension answers given by their students. The discussions should not focus on what are the right answer(s) but on how to arrive at the right answer(s).

What is important from such discussions for reading teachers is to understand how the students create meaning from text. The CRITEXT information gathered from the discussions will be useful in helping poor readers to develop a basis for their own understanding and effective use of texts. CRITEXT may also be useful in guiding reading teachers to identify comprehension problems faced by poor comprehenders, especially in comprehension questions that require contextual critical analysis. Related to CRITEXT is the notion of promoting critical reading in the EFL or ESL classroom. Wallace (1992, p. 61) says 'Critical reading has not been generally encouraged in the English as a Foreign Language (EFL) classroom ... [and] Students tend not to be invited to draw on their own experiences of literacy, or to articulate their understanding of it as a social phenomenon.' Thus, it is crucial to develop critical awareness on the part of the readers in the reading classroom and CRITEXT may serve this purpose.

In conclusion, what this contextual critical analysis demands is that the depth of an answer needs to be fine-tuned to the requirements of the question and the context of the text. Appropriate critical reasoning in question answering strategies, which can be learned from Dole et al. (1991) and the DARTs in Lunzer et al. (1984), can be further practised and developed by reading teachers through understanding how readers reach their selected answers. This means that the

issue of the acceptability of the reasoning process in reading comprehension needs to be further developed, especially in the context of the unitariness of the eight subskills in comprehension answering strategies in both languages. This critical awareness in understanding the information in the text requires a shift from total emphasis on the product approach to a more active process orientation.

BIBLIOGRAPHY

- Adams, M. J. (1993) 'Beginning to read: an overview'. In R. Beard (ed.) *Teaching Literacy: Balancing Perspectives*. London: Hodder and Stoughton.
- Alderson, J.C. (1984) 'Reading in a foreign language: A reading problem or a language problem? '. In J. C. Alderson and A. H. Urquhart (eds) *Reading in a Foreign Language* . (pp. 1-27). New York: Longman.
- Alderson, J. C. (1990) 'Testing reading comprehension skills (part one)'. *Reading in a Foreign Language*, 6, 2, 425-438.
- Anderson, R. C. and Pearson, P. D. (1984) 'A schema-theoretic view of basic processes in reading'. In P. D. Pearson (ed.) *Handbook of Reading Research*. (pp. 255-292). New York: Longman.
- Atan bin Long *et al.* (1984) *A Survey of the Reading Habits and Interests of Malaysian People*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Auerbach, E. R. (1989) 'Toward a socio-contextual approach to family literacy'. *Harvard Educational Review*, 59, 2, 165-181.
- Baker, L. and Brown, A. L. (1984) 'Metacognitive skills and reading'. In P. D. Pearson (ed.) *Handbook of Reading Research* (pp. 353-394). New York: Longman.
- Beard, R. (ed.) (1993) *Teaching Literacy: Balancing Perspectives*. London: Hodder and Stoughton.
- Beard, R. (1990) *Developing Reading 3-13* (2nd edn). London: Hodder and Stoughton.

Bernhardt, E. B. (1991) *Reading Development in a Second Language: Theoretical, Empirical, and Classroom Perspectives*.
New Jersey: Ablex Publishing Corporation.

Bernhardt, E. B. (1991) 'A psycholinguistic perspective on second language literacy'. In J. H. Hulstijn and J. F. Matter (eds) *Reading in Two Languages*. AILA Review, 8, 31-60.

Block, E. L. (1992) 'See how they read: comprehension monitoring of L1 and L2 readers'. *TESOL Quarterly*, 26, 2, 319-343.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D.R. (1956) *Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain*. New York: Longmans, Green.

Brannen, J. (ed.) (1992) *Mixing Methods: Qualitative and Quantitative Research*. Aldershot: Avebury.

Brooks, G. (1984) 'Nineteenth- and twentieth-century models of L2 reading'. In A. K. Pugh and J. M. Ulijn, (eds) *Reading for Professional Purposes: Studies and Practices in Native and Foreign Languages*.
London: Heinemann.

Brown, A. L. and Palincsar, A. S. (1989) 'Guided, cooperative learning and individual knowledge acquisition'. In L. B. Resnick (ed.) *Knowing, Learning and Instruction. Essays in Honour of Robert Galser*. Hillsdale, NJ: Erlbaum. (pp. 393-451)

Brown, J. D. (1988) *Understanding Research in Second Language Learning*.
Cambridge: Cambridge University Press.

Bryant, P. E., Bradley, L., Maclean, M. and Crossland, J. (1989) 'Nursery rhymes, phonological skills and reading'. *Journal of Child Language*, 16, 407-428.

Bryman, A. and Cramer, D. (1990) *Quantitative Data Analysis for Social Scientists*. London: Routledge.

Carrell, P. L. (1987) 'A view of written text as communicative interaction: implications for reading in a second language'. In J. Devine, P. L. Carrell and D. E. Eskey (eds) *Research in Reading in English as a Second Language*. Washington, D.C. : TESOL

Carroll, J. B. (1978) 'The nature of the reading process'. In L. J. Chapman and P. Czerniewska (eds) *Reading From Process to Practice*. London: Routledge & Kegan Paul & The Open University Press.

Casanave, C. P. (1988) 'Comprehension monitoring in ESL reading: a neglected essential'. *TESOL Quarterly*, 22 (2), 283-302.

Chall, J. S. (1983) *Learning to Read: The Great Debate* (2nd updated edn). New York: McGraw-Hill.

Chall, J. S. (1977) *Reading 1967-1977: A Decade of Change and Promise*. Bloomington, Indiana, USA: The Phi Delta Kappa Educational Foundation.

Chall, J. S. (1983) *Stages of Reading Development*. New York: McGraw-Hill.

Chapman, L. J. (1987) *Reading: From 5-11 Years*. Milton Keynes: Open University Press.

Chapman, L. J. and Czerniewska, P. (eds) (1978) *Reading: From Process to Practice*. London: Routledge and Kegan Paul and Open University Press.

Child, D. (1970) *The Essentials of Factor Analysis*. London: Holt, Rinehart and Winston.

- Clark, H. (1975) 'Comprehension as an art' (Unpublished manuscript). In K. Kavale, and R. Schreiner (1979) 'The reading processes of above average and average readers: a comparison of the use of reasoning strategies in responding to standardized comprehension measures'. *Reading Research Quarterly* , 15, 1, 102-128.
- Clarke, M., and Silberstein, S. (1977) 'Toward a realization of psycholinguistic principles for the ESL reading class'. *Language Learning*, 27, 135-154.
- Clay, M. M. (1979) *The Early Detection of Reading Difficulties: A Diagnostic Survey with Recovery Procedures* (2nd edn). Auckland: Heinemann.
- Clegg, F. (1990). *Simple Statistics*. Cambridge: Cambridge University Press.
- 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*. pp. 5-12. Rowley, MA: Newbury House.
- Cohen, L. and Manion, L. (1989) *Research Methods in Education* (3rd. edn). London: Routledge.
- Dale, P. S. (1976) *Language Development: Structure and Function*. New York: Holt Rinehart & Winston.
- Davies, F. (1995) *Introducing Reading*. London: Penguin.
- Davies, F. and T. Greene (1982) Directed Activities Related to Text: Text analysis and text reconstruction. *ERIC Clearing House on Communication Skills and Reading, Abstract in Resources in Education*, March 1982.
- Davies, F and Greene, T. (1984) *Reading for Learning in the Sciences*. Edinburgh: Oliver and Boyd.

- Davis, F. B. (1944) 'Fundamental Factors of Comprehension in Reading'. *Psychometrika*, 9, 185-197.
- Davis, F. B. (1946) 'A brief comment on Thurstone's note on a reanalysis of Davis' reading tests'. *Psychometrika*, 11,4, 249-255.
- Davis, F. B. (1968) 'Research in comprehension in reading'. *Reading Research Quarterly*, 3, 499-545.
- Davis, F. B. (ed.) (1971) *The Literature of Research in Reading with Emphasis on Models*. Final Report, US Department of Health, Education and Welfare.
- Davis, F. B. (1972) 'Psychometric research on comprehension in reading'. *Reading Research Quarterly*, 7, 628-678.
- Day, R. R., Omura, C. and Hiramatsu, M. (1991) 'Incidental EFL vocabulary learning and reading'. *Reading in a Foreign Language*, 7, 2, 541-551.
- Devine, J., Carrell, P. L. and Eskey, D. E. (eds.) (1987) *Research in Reading in English as a Second Language*. Washington, D.C.: TESOL
- Dole, J. A., Duffy, G. G., Roehler, L. R. and Pearson, P. D. (1991) 'Moving from the old to the new: research on reading comprehension instruction'. *Review of Educational Research*, 61, 2, 239-264.
- Dombey, H. (1993) 'Reading: what children need to learn and how teachers can help them'. *Reading*, 27, 3, 2-19.
- Downing, J. (1982) 'Reading- skill or skills?'. *The Reading Teacher*, 35, 534-537.
- Duffy, G. G., Roehler, L. R., Sivan, E. *et al.* (1987). 'Effects of explaining the reasoning associated with using reading strategies'. *Reading Research Quarterly*, 22, 347-368.

Faerch, C. and Kasper, G. (eds) (1987) *Introspection in Second Language Research*. Clevedon, UK: Multilingual Matters.

Farr, R. (1969) *Reading: What can be measured?* IRA. Delaware.

Flavell, J. M. (1977) *Cognitive Development*. New York: Prentice- Hall.

Flesch, R. (1955) *Why Johnny Can't Read and What You Can Do About It*. New York: Harper and Row.

Fyfe, R. and Mitchell, E. (1985) *Reading Strategies and Their Assessment*. Berkshire: NFER-NELSON.

Golinkoff, R. (1976) 'A comparison of reading comprehension processes in good and poor Comprehenders'. *Reading Research Quarterly*, 11,4,623-659.

Goodman, K. S. (1967) 'Reading: a psycholinguistic guessing game'. *Journal of the Reading Specialist*, 6, 125-135.

Goodman, K. S. (ed.) (1968) *The Psycholinguistic Nature of the Reading Process*. Detroit, MI: Wayne State University Press.

Goodman, K. S. (1970) 'Reading: a psycholinguistic guessing game'. In H. Singer and R. B. Ruddell (eds) *Theoretical Models and Processes of Reading*. Newark, DE: International Reading Association.

Goodman, K. S. (1982) 'What is basic about reading?'. In F. V. Gollasch (ed.) *Language and Literacy: The Selected Writings of Kenneth S. Goodman*. London: Routledge and Kegan Paul.

Gough, P. B. (1976) 'One second of reading'. In H. Singer and R. B. Ruddell (eds) *Theoretical Models and Processes of Reading*. (2nd edn). Newark, DE: International Reading Association.

Grabe, W. (1991) 'Current developments in second language reading research'. *TESOL Quarterly*, 25, 3, 375-406.

Grellet, F. (1981) *Developing Reading Skills: A Practical Guide to Reading Comprehension*. Cambridge: Cambridge University Press.

Haarstrup, K. (1987) 'Using think-aloud and retrospection to uncover learner's lexical inferencing procedures'. In C. Faerch and G. Kasper (eds) (1987) *Introspection in Second Language Research*. Clevedon, UK: Multilingual Matters.

Haji Husin, K. (1992) *KBSM dan Model-Model Pengajaran Bahasa* [KBSM and Models for Language Teaching]. (3rd edn). Kuala Lumpur: Kumpulan Budiman.

Haji Omar, A. (1992) *The Linguistic Scenery in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Haji Sabran, R. & Syam, R. (1985) *Kaedah Pengajaran Bahasa Malaysia*. [The Teaching Methodologies of Bahasa Malaysia]. Petaling Jaya: Fajar Bakti.

Harrison, C. (1979) 'Assessing the readability of school texts'. In E. Lunzer and K. Gardner (eds) *The Effective Use of Reading*. London: Heinemann Educational Books for The Schools Council.

Harrison, C. (1992a) 'The reading process and learning to read: what a teacher using a 'real books' approach needs to know'. In C. Harrison and M. Coles (eds) *The Reading for Real Handbook*. London: Routledge Kegan Paul.

Harrison, C. (1992b) 'Information skills'. In F. Satow and B. Gatherer (eds) *Literacy Without Frontiers*. United Kingdom Reading Association.

- Harrison, C. (1994) 'The assessment of response to reading: developing a postmodern perspective'. In A. Goodwyn (ed.) *English and Ability*. London: David Fulton.
- Harrison, C. (1996) *The Teaching of Reading: What Teachers Need to Know*. United Kingdom: UKRA
- Hatch, E. and Lazaraton, A. (1991) *The Research Manual: Design and Statistics for Applied Linguistics*. New York: Newbury House.
- Hittleman, D. R. (1978) *Developmental Reading: A Psycholinguistic Perspective*. Chicago: Randy McNally.
- Hitchcock, G. and Hughes, D. (1989) *Research and the Teacher: A Qualitative Introduction to School-Based Research*. London: Routledge.
- Hosenfeld, C. (1977) 'A preliminary investigation of the reading strategies of successful and unsuccessful second language learners'. *System*, 5, 110-23.
- Hulstijn, J. H. (1991) 'How is reading in a second language related to reading in a first language?'. In J. H. Hulstijn and J. F. Matter (eds) *Reading in Two Languages (AILA Review)*, 8, 5-14.
- Jenkinson, M. D. (1978) 'Ways in teaching'. In L. J. Chapman and P. Czerniewska (eds) *Reading From Process to Practice*. London: Routledge & Kegan Paul & The Open University Press.
- Johnston, P. H. (1983) *Reading Comprehension Assessment: A Cognitive Basis*. Newark, Del: International Reading Association.
- Johnston, P. H. (1984) 'Assessment in Reading'. In P. D. Pearson (ed.) *Handbook of Reading Research*. New York: Longman.

- Kamil, M. K. (1984) 'Current traditions of reading research'. In P. D. Pearson (ed.) *Handbook of Reading Research* (39-62). New York: Longman.
- Kavale, K. and Schreiner, R. (1979) 'The reading processes of above average and average readers: a comparison of the use of reasoning strategies in responding to standardized comprehension measures'. *Reading Research Quarterly* , 15, 1, 102-128.
- Kementerian Pendidikan Malaysia (1982) Sukatan Pelajaran Sekolah Rendah 1983 : Bahasa Inggeris [Ministry of Education, Malaysia. The Syllabus Content for Primary School 1983: English Language]. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kementerian Pendidikan Malaysia (1988) Sukatan Pelajaran Sekolah Menengah 1988: Bahasa Inggeris. [Ministry of Education, Malaysia. The Syllabus Content for Secondary School 1988: English Language]. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kementerian Pendidikan Malaysia (1989a) *Compendium: A Concise Guide For Teachers of English. Committee to Plan and Co-ordinate English Language Programmes in Schools*. Kuala Lumpur: Malaysian Ministry of Education.
- Kementerian Pendidikan Malaysia (1989b) Huraian Sukatan Pelajaran Bahasa Inggeris Tingkatan III Kurikulum Bersepadu Sekolah Menengah. [Ministry of Education, Malaysia. An Illustration of the Syllabus Content for English Language Form III Integrated Curriculum for the Secondary School]. Kuala Lumpur: Kementerian Pendidikan Malaysia.

Kementerian Pendidikan Malaysia (1989c) Huraian Sukatan Pelajaran Bahasa Malaysia Tingkatan III Kurikulum Bersepadu Sekolah Menengah.

[Ministry of Education, Malaysia. An Illustration of the Syllabus Content for Bahasa Malaysia Form III Integrated Curriculum for the Secondary School]. Kuala Lumpur: Kementerian Pendidikan Malaysia.

Kementerian Pendidikan Malaysia (1990) Buku Panduan Khas Bahasa Malaysia Tahun Enam Kurikulum Baru Sekolah Rendah. [Ministry of Education, Malaysia. Special Guide Book for Bahasa Malaysia for Year Six. New Curriculum for Primary School]. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kementerian Pendidikan Malaysia (1995) Sukatan Pelajaran Sekolah Rendah: Bahasa Melayu. [Ministry of Education, Malaysia. Syllabus Content for Primary School: Malay Language]. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kletzien, S. B. (1991) 'Strategy use by good and poor comprehenders reading expository text of differing levels'. *Reading Research Quarterly* . 26, 1, 67-86.

Kline, P. (1994) *An Easy Guide to Factor Analysis*. London: Routledge.

Lo, S. M. R. (1992) *A Study of the Comprehension Processes of First and Second Language Learners*. Ph.D Thesis (Unpublished), The University of Nottingham, School of Education.

Lunzer, E. and Gardner, K. (eds) (1979) *The Effective Use of Reading*. London: Heinemann Educational Books for The Schools Council.

Lunzer, E., Gardner, K., Davies, F. and Greene, T. (1984) *Learning From the Written Word*. Schools Council Publications. Edinburgh: Oliver and Boyd.

- Lunzer, E., Waite, M. and Dolan, T. (1979) 'Comprehension and comprehension tests'. In E. Lunzer, and K. Gardner (eds) *The Effective Use of Reading*. London: Heinemann Educational Books for The School Council.
- Marshall, C. and Rossman, G. B. (1989) *Designing Qualitative Research*. Newbury Park: SAGE Publications.
- McDonough, S. H. (1995) *Strategy and Skill in Learning a Foreign Language*. London: Edward Arnold.
- Miller, G. A. (1963) *Language and Communication*. New York: McGraw-Hill.
- Mulaik, S. A. (1972) *The Foundations of Factor Analysis*. New York: McGraw-Hill Company.
- Mustapha, N. F. (1994) Reading and writing research: the Malaysian experience. Paper presented at the International Seminar on Reading and Writing Research: Implications for Language Education, SEAMEO, Regional Language Centre, Singapore, 18th-20th April.
- Myers, S. S. (1991) 'Performance in reading comprehension-product or process?'. *Educational Review*, 43, 3, 257-272.
- Newell, A. and Simon, H. A. (1972) *Human Problem Solving*. Englewood Cliffs, N. J.: Prentice Hall.
- Newmark, P. (1989) 'Introductory survey'. In C. Picken (ed.) *The Translator's Handbook*. (2nd Edn). London: Aslib.
- Nicholson, T. (1993) 'Reading without context'. In G. B. Thompson, W. E. Tunmer, and T. Nicholson (eds) *Reading Acquisition Processes*. Clevedon: Multilingual Matters.

- Oakhill, J. (1993) 'Developing skilled reading'. In Beard, R. (ed.) *Teaching Literacy: Balancing Perspectives*. London: Hodder and Stoughton.
- Olshavsky, J. E. (1976-1977) 'Reading as problem-solving: An investigation of strategies'. *Reading Research Quarterly*, 12,4, 654-674.
- Olshavsky, J. E. (1978) 'Comprehension profiles of good and poor readers across materials of increasing difficulty'. In P. D. Pearson and J. Hansen (eds) *Reading: Disciplined Inquiry in Process and Practice. Twenty eight yearbook of the National Reading Conference*. (pp.73-76). Clemson, SC: National Reading Conference.
- Paris, S. G., Lipson, M. Y. and Wixson, K. K. (1994) 'Becoming a Strategic Reader'. In R. B. Ruddell, M. R. Ruddell and H. Singer (eds) *Theoretical Models and Processes of Reading*. (4th edn). Newark, DE International Reading Association.
- Pearson, P. D. and Stephens, D. (1994) 'Learning about literacy: A 30-year journey '. In R. B. Ruddell, M. R. Ruddell and H. Singer (eds) *Theoretical Models and Processes of Reading*. (4th edn). Newark, DE International Reading Association.
- Pressley, M., Johnson, C. J. *et al.* (1989) 'Strategies that improve children's memory and comprehension of texts'. *The Elementary School Journal*, 90, 3-32.
- Pritchard, R. (1990) 'The effects of cultural schemata on reading processing strategies'. *Reading Research Quarterly*, 25, 4, 273-91.
- Ramaiah, M. (1994) 'Reading initiatives in Malaysia'. In M. L. Tickoo (ed.) *Research in Reading and Writing: A Southeast Asian Collection*. Anthology Series 32. Singapore: SEAMEO Regional Language Centre.

- Rayner, K. and Pollatsek, A. (1989) *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice Hall.
- Robson, C. (1993) *Real World Research*. Oxford: Blackwell.
- Rosenshine, B. V. (1980) 'Skill hierarchies in reading comprehension'. In R. J. Spiro, B. C. Bruce and W. F. Brewer (eds) *Theoretical Issues in Reading Comprehension*. New Jersey: LEA.
- Rost, D. H. (1989) 'Reading comprehension: skill or skills?'. *Journal of Research in Reading*, 12, 2, 87-113.
- Ruddell, M. R. (1994) 'Vocabulary knowledge and comprehension: a comprehension-process view of complex literacy relationships'. In R. B. Ruddell, M. R. Ruddell and H. Singer (eds) *Theoretical Models and Processes of Reading*. (4th edn). (pp. 414-447) Delaware: IRA.
- Ruddell, R. B. and Ruddell, M. R. (1994) 'Language acquisition and literacy processes'. In R. B. Ruddell, M. R. Ruddell and H. Singer (eds) *Theoretical Models and Processes of Reading*. (4th edn) (pp. 83-103) Delaware: IRA.
- Rumelhart, D. E., and Ortony, A. (1977) 'The representation of knowledge in memory '. In R. C. Anderson, R. J. Spiro and W. E. Montague (eds) *Schooling and the Acquisition of Knowledge*. (pp. 99-136). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Safiah Osman (1985) 'The effect of content familiarity of the comprehension of text by learners of English as a Second Language'. In Siew Tin Bennett and Zawiah Yahya (eds) *Proceedings of the National Seminar on Recent Developments in Materials and Classroom Techniques in TESL*. UKM: Pusat Bahasa. 32-34.

- Samuels, S. J. (1980) 'The age-old controversy between holistic and subskills approaches to beginning reading instruction revisited'. In C. M. McCullough, (ed.) *Inchworm, Inchworm: Persistent Problems in Reading Education*. Delaware: IRA.
- Samuels, S. J. and Kamil, M. K. (1984) 'Models of the reading process'. In P. D. Pearson (ed.) *Handbook of Reading Research*. (pp. 185-224). New York: Longman.
- Sarig, G. (1987) 'High-level reading in the first and foreign language: some comparative process data'. In Devine, J. et al. (eds) *Research in Reading English as a Second Language*. Rowley, Mass.: Newbury House.
- Sarig, G. (1988) Comprehension-promotion strategies: the sum of the parts and the whole. Paper presented at the Second International Congress of Applied Psycholinguistics. University of Kassel.
- Schmitt, M. C. and Baumann, J. F. (1990) ' Metacomprehension during basal reader instruction: do teachers promote it? '. *Reading Research and Instruction*, 29, 2-13.
- Scott, M. (1990) *Demystifying the Jabberwocky: A Research Narrative*. Unpublished Ph.d. thesis, University of Lancaster.
- Smith, F. (1971) *Understanding Reading*. New York: Holt, Rinehart and Winston.
- Smith, F. (1973) (ed.) *Psycholinguistics and Reading*. New York: Holt, Rinehart and Winston.
- Spache, G.D. and Spache, E.B. (1969) *Reading in the Elementary School*. (2nd edn) Boston: Allyn and Bacon.

- Spearritt, D. (1972) 'Identification of subskills in reading comprehension by maximum likelihood factor analysis'. *Reading Research Quarterly*, 8, 92-111.
- Spiro, R. J. (1980) 'Constructive processes in prose comprehension and recall'. In R. J. Spiro, B. C. Bruce and W. F. Brewer (eds) *Theoretical Issues in Reading Comprehension* (pp. 245-278). Hillsdale, N. J. : Erlbaum.
- Stanovich, K. (1980) 'Toward an interactive-compensatory model of individual differences in the development of reading fluency'. *Reading Research Quarterly*, 16, 1, 32-71.
- Stanovich, K. (1986) 'Matthew effects in reading: some consequences of individual differences in the acquisition of literacy'. *Reading Research Quarterly*, 21, 360-407.
- Stedman, L. C. and Kaestle, C. F. (1987) 'Literacy and reading performances in the United States'. *Reading Research Quarterly*, 22, 8-46.
- Steffensen, M.S. and Joag-Dev, C. (1984) 'Cultural knowledge and reading'. In J. C. Alderson and A. S. Urquhart (eds) *Reading in a Foreign Language*. Harlow: Longman, 48-61.
- Steffensen, M.S. (1987) 'The effects of context and culture on children's L2 reading: a review'. In J. Devine, P. L. Carrell and D. E. Eskey, (eds) *Research in Reading in English as a Second Language*. Washington, D.C. : TESOL
- Steiner, G. (1975) *After Babel: Aspects of Language and Translation*. London: Oxford University Press.
- Stierer, B. and Bloome, D. (1994) *Reading Words: A Commentary on the Key Terms in the Teaching of Reading*. Sheffield: NATE.

- Tabachnick, B. G. and Fidell, L. S. (1989) *Using Multivariate Statistics*. New York: Harper Collins.
- Thorndike, R. B. (1973) *Reading Comprehension Education in Fifteen Countries*. (International Studies in Evaluation III) New York: John Wiley.
- Thorndike, E. L. (1971) 'Reading as reasoning: a study of mistakes in paragraph reading'. *Reading Research Quarterly*, 6, 425-434.
- Thurstone, L.L. (1946) 'Note on a reanalysis of Davis' reading tests'. *Psychometrika*, 11 (3), 185-188.
- Thurstone, L.L. (1947) *Multiple Factor Analysis: A Development and Expansion of Vectors of the Mind*. Chicago: University of Chicago Press.
- Tierney, R. J. and Cunningham, J. W. (1984) 'Research on teaching reading comprehension'. In P. D. Pearson (ed.) *Handbook of Reading Research*. London: Longman.
- Tierney, R. J. and Pearson, P. D. (1994) 'Learning to learn from text: a framework for improving classroom practice'. In R. B. Ruddell, M. R. Ruddell and H. Singer (eds) *Theoretical Models and Processes of Reading*. (4th edn) (pp. 496-513) Delaware: IRA.
- Tuckman, Bruce W. (1978) *Conducting Educational Research*. (2nd edn). New York: Harcourt Brace Jovanovich.
- Turner, M. (1990) *Sponsored Reading Failure*. IPSET Education Unit.
- Valencia, S. W. and Pearson, P. D. (1988) 'Principles for classroom comprehension assessment'. *Remedial and Special Education (RASE)*, 9, 1, 26-35.
- Venezky, Richard L. (1984) 'The history of reading research'. In P. D. Pearson (ed.) *Handbook of Reading Research* (pp. 3-38). New York: Longman.

- Wagoner, S. A. (1983) 'Comprehension monitoring: what it is and what we know about it'. *Reading Research Quarterly*, 18, 3, 328-346.
- Waite, M. (1980) An Attempt to Identify and Measure Certain Subskills of Reading Comprehension. Unpublished M. Phil. Thesis, The University of Nottingham, School of Education.
- Wallace, C. (1992) *Reading*. Oxford: Oxford University Press.
- Wallace, C. (1992) 'Critical literacy awareness in the EFL classroom'. In N. Fairclough (ed.) *Critical Language Awareness* (pp. 59-92). London: Longman
- Waterland, L. (1988) *Read With Me: An Apprenticeship Approach to Reading* (2nd edn). Stroud: Thimble Press.
- Weaver, C. (1990) *Understanding Whole Language: From Principles to Practice*. Portsmouth, N. H. : Heinemann.
- Weiss, N. A. (1995). *Introductory Statistics*. (4th edn). Reading, Massachusetts: Addison-Wesley.
- Wright, G. & Fowler, C. (1986) *Investigative Design and Statistics*. Middlesex: Penguin.
- Willinsky, J. (1990) *The New Literacy: Redefining Reading and Writing in the Schools*. New York: Routledge.
- Youngman, M. B. (1979) *Analysing Social and Educational Research Data*. London: McGraw-Hill.
- Yin, R. K. (1994) *Case Study Research: Design and Methods*. London: Sage Publications.

