CHAPTER II

LITERATURE REVIEW

Chapter Two presents a review of literature on theoretical perspectives which are pertinent to the areas of reading comprehension, language testing, test-takers' computer related variables, and the research approach and statistics. This chapter begins with a discussion of studies on two major dimensions of reading comprehension which are the reader and the text. Then studies on three areas of language testing which are test characteristics, reading tests, and computerized tests are explored. Subsequently, this chapter focuses on the test-takers' variables and literature that is related to the three selected variables, the relationships among themselves, and their impact on performances. The final part of this chapter includes a review of the studies on correlational research approach and statistics focused on correlational studies.

Reading Comprehension

In Longman Dictionary of Language Teaching & Applied Linguistics, Richards, Platt, and Platt (1999: 306-307) defines the term "reading" as "perceiving a written text in order to understand its contents." The written text can be read silently (silent reading). The understanding that results is called "reading comprehension." Alternatively, the written text can also be read by saying the text aloud (oral reading). This can be done with or without understanding of the contents. This part of the literature review focuses on silent reading only because it is used in the reading comprehension test of this study.

From the given definitions mentioned above, two major factors are directly concerned with reading comprehension. They are the reader and the text. The following sections discuss these two factors in detail.

The Reader

This part explores those studies that consider theoretical aspects of an individual person that assumes the role of a reader. The studies include communicative competence, the nature of reading ability, the first language readers and the second/foreign language readers, and the schema theory.

1. Communicative Competence

Recently, the role of communicative competence has been increasingly emphasized among the language teaching and assessment community. The idea of communicative competence addresses "language use" as a primary component in language teaching and assessment. The concept was proposed by Hymes (1972) and Campbell and Wales (1970). This perspective is an alternative to the one proposed by Chomsky that linguistic competence was limited to the knowledge of grammatical rules. Communicative competence views the role played by context of discourse beyond sentential constructions. The current models include not only the knowledge of grammar, but also both the knowledge of language functions and the knowledge of language contexts. One of the well-known frameworks is proposed by Canale and Swain (1980) which includes grammatical competence, sociolinguistic competence, and strategic competence.

In 1990, Bachman proposed a communicative framework which includes grammatical competence and textual competence, organized under organizational competence with illocutionary competence and sociolinguistic competence under pragmatic competence. This framework is important in language assessment in that it offers some mechanism for explaining the extent to which an individual's language performance might vary across tasks. Besides the trait factors of linguistic competence, pragmatic competence, and strategic competence, Bachman's framework also adds method factors which are the channel, the mode, the language-use situation, the distribution of information, the type of information, etc. Whereas the competence is about general language abilities, the method relates the abilities to the real-world language performance which might help address the competence versus performance issue (Skehan, 1991).

Therefore, language assessment with a communicative competence perspective "must be devoted not only to what the learner knows about the second language and about how to use it (competence) but also to what extent the learner is able to actually demonstrate this knowledge in a meaningful communicative situation" (Canale and Swain, 1980: 32).

2. The Nature of Reading Abilities

To describe the nature of reading abilities, four important components need to be included. The first component describes the purposes of reading given that each purpose emphasizes to a certain extent different combinations of skills and strategies. The second component emphasizes many criteria that define the nature of fluent reading abilities. The third component addresses how reading is carried out as a cognitive process and the last component explains the ability to draw meaning from a text and interpret this meaning. This ability varies in line with the second language proficiency of the reader (Grabe and Stoller, 2002: 9-10).

2.1 Purposes of Reading

Anderson (1999) mentioned that readers move through the printed text with specific purposes in mind to accomplish specific goals.

Grabe and Stoller (2002: 12) classified the purposes of reading into seven categories.

1. Reading to search for simple information
2. Reading to skim quickly
3. Reading to learn from texts
4. Reading to integrate information
5. Reading to write
6. Reading to critique texts
7. Reading for general comprehension

Reading to search for simple information is a process of scanning the text for a specific word or information. Reading to skim involves a combination of strategies for guessing where important information might be in the text, then using basic reading comprehension skills on that part of the text until a general idea is found. Reading to learn occurs in academic and professional contexts.

Normally it is carried out at a slower reading rate and it requires stronger inferential skill to connect text information with background knowledge. Reading to integrate information requires additional decisions about the multiple sources of text and requires critical evaluation of the information from the text. Reading to write and reading to critique texts are different tasks of reading to integrate information. Finally, reading for general comprehension is the most basic purpose for reading but it is more complex than commonly assumed. When accomplished by a skilled fluent reader, it requires very rapid and automatic processing of words, strong skills in forming a general meaning representation of main ideas, and efficient coordination of many processes under very limited time constraints (Grabe and Stoller, 2002: 13-14).

Grabe and Stoller (2002: 15) identified "skills" as the "linguistic processing abilities that are relatively automatic in their use and their combinations e.g., word recognition, syntactic processing". They also mentioned that in general "strategies" are often defined as "a set of abilities under conscious control of the reader". Nevertheless, they mentioned that many abilities that are identified as strategies are relatively automatic in their use by fluent readers e.g., skipping an unknown word while reading. There is, thus, no clear distinction between skills and strategies. However, for definitional purposes Grabe and Stoller (2002: 17) defined "strategies" as "abilities that are potentially open to conscious reflection and use". They use the term "abilities" to refer to what readers do while reading and "processes" to refer to mental operations.

Alderson (2000a) defined the "process" of reading as the interaction between a reader and the text. He explained that while reading, the reader is presumably also thinking about what is being read, what it means to the reader, how it relates to other things that the reader has read, to the things that the reader knows, to what the reader expects to come next in the text. The reader presumably also thinks about how useful, entertaining, or boring the text is. Alderson (2000a) mentioned that evidence indicated that many different things happen when a reader reads. The process seems to be dynamic, variable and different for a different reader on a different text at a different time or with a different purpose of reading that. Alderson (2000a) suggested that understanding the reading process is important to the

understanding of the nature of reading. However, to understand the reading process is certainly a difficult thing to do because the process is silent, internal, and private.

Alderson (2000a) explained that an alternative to the process of reading is to inspect the "product" of reading. Although different readers may have different reading processes to reach a given understanding, what matters is what understanding you do reach. Earlier in this century, many scholars and researchers focused on the product approaches to reading and investigated what the reader has understood after reading a particular text. Bloom's (1956) taxonomy of reading skills, Davis's (1968) eight reading skills, and Munby's taxonomy (1978) are examples of reading experts who identified reading in terms of a list of readers' abilities. However, the use of taxonomies was criticized because some of them do not have empirical data to support the claims and others do not have real discrete definitions (Alderson, 2000a).

On the issue of divisibility of language skills, Bae and Bachman (1998) mentioned that experts in the field of reading agree that language ability consists of multiple aspects, components, or competencies. However, questions still remain regarding what these specific components are, and more importantly, the extent to which language tests are capable of differentiating among them. Alderson (2000a) mentioned that in recent years the product approaches to reading have been unfashionable and researchers have concentrated on understanding the reading process. Some prominent studies in this perspective are reviewed and discussed in the next three sections.

2.2 Definition of Fluent Reading Comprehension

The term "reading comprehension" as defined earlier is the ability to understand the contents or information in a text and properly interpret it. Nevertheless, reading comprehension abilities are far more complex than the definition suggests. Grabe and Stoller (2002: 17) define fluent reading comprehension in terms of "the processes" and they include 10 processes in their definition.

1. A rapid process
2. An efficient process
3. An interactive process

4. A strategic process
5. A flexible process
6. An evaluating process
7. A purposeful process
8. A comprehending process
9. A learning process
10. A linguistic process

Fluent readers read **rapidly** at the rate of between 200 and 300 words per minute when the reading process is carried out **efficiently**. Reading is an **interactive** process in two ways. Firstly, various reading processes are carried out simultaneously, such as analyzing sentence structure, building the main idea, and monitoring comprehension. Secondly, it is interactive in the sense that linguistic information from the text interacts with the background knowledge of the reader.

Readers need to be **strategic** to be able to balance the many skills needed for comprehension. At the same time, readers need to be **flexible** enough to get in line with changing purposes and the ongoing monitoring of comprehension. Reading is an **evaluating** process in that readers must decide if the information is coherent and matches the reading purpose.

Reading is always **purposeful**. Readers read in different ways, not only based on their different purposes, but also in the sense that any internally or externally imposed motivation is triggered by some individual purpose or task. Reading is a **comprehending** process which is obvious in that understanding a text is the purpose of reading. And one outcome of reading is a **learning** process where readers learn new information. Finally, reading is a **linguistic** process as the reader has to understand words to comprehend the text.

Thus, reading comprehension abilities are quite complex and involve different processes in different ways depending on the goals, the motivations, and the language ability of the readers. However, a set of common underlying processes are activated while we are reading. In the next section, this set of common underlying process which seems to work for skilled readers is explored.

2.3 Components of Reading Abilities

Grabe and Stoller (2002: 19) mentioned that as we read there is a set of common underlying purposes that are activated. Assuming a purpose of general comprehension of a longer text, they divided the reading comprehension processes of fluent readers into lower-level processes and higher-level processes as follows:

Lower-level processes include

- Lexical access
- Syntactic parsing
- Semantic proposition formation
- Working memory activation.

Higher-level processes deal with

- Text model of comprehension
- Situation model of reader interpretation
- Background knowledge use and inferencing
- Executive control processes.

Grabe and Stoller (2002: 20-24) explained that the lower-level processes are automatic and skills orientated. The most basic requirement for the fluent reader is rapid and automatic word recognition or **lexical access**. The fluent reader is also able to take in and store words together so that basic grammatical information can be extracted. This process is called **syntactic parsing**. The third basic and automatic process is the process of combining word meanings and structural information into basic clause-level meaning units and this process is called **semantic proposition formation**. The three processes work together effortlessly in working memory. It is the network of information and related processes that are being used at a given moment which is **working memory activation**. When the words are accessed, the information is cued grammatically, and the emerging meanings are all active for a short period of time in the working memory which keeps information active for one to two seconds while it carries out the appropriate processes. Speed of processing is essential because without the rapid process of active information, the information might fade from the memory and would then need to be reactivized. In this situation the reading process would not be efficient.

Grabe and Stoller (2002: 25-29) also stated that higher-level processes represent comprehension processes that make much more use of the reader's background knowledge in inferencing skills. The most basic higher-level comprehension process is the coordination of ideas from a text that represents the main points and supporting ideas to form a meaning representation of the text which is called a **text model of reading comprehension**. When readers read, some new information is connected to their network of ideas. If those ideas are used repeatedly, they will remain in the network. Other ideas that do not have any role in connecting new information tend to be removed from the network. If the text information is completely new and difficult to understand or the readers have limited language proficiency which obstructs the comprehension, then the inferencing skills are extensively required by readers. At this point, **background knowledge** takes a supporting role by helping the reader find the meaning of the ambiguous words and clauses and anticipate the discourse organization of the text as new information is incorporated into the text model of comprehension. At the same time that the text model of comprehension is being built, the reader almost immediately begins to interpret information from the text in terms of the reader's own goals, feelings, and background knowledge. This process is called the **situation model of reader interpretation**. The construction of the text model and the situation model require the abilities to monitor comprehension, use strategies as needed, reassess and reestablish goals, and repair comprehension problems. However, it is not clearly understood how such a monitor (**executive control processing**) might operate cognitively.

In summary, the reading comprehension processes incorporate and balance many abilities in a very complex and rapid set of routines that make reading comprehension an effortless and pleasant experience for fluent readers. Lack of linguistic resources or background knowledge can lead to reading difficulties.

2.4 Models of Reading

Researchers of reading skills and processes attempt to create a reasonable framework to understand the reading comprehension process and to do so they come up with different models of reading. Grabe and Stoller (2002: 31) categorized models of reading into two groups. The general models of reading serve

useful purposes, mostly by providing a metaphorical interpretation of the processes involved in reading comprehension. The other group models of reading are more specific in nature, trying to account for and interpret the results of many research studies.

2.4.1 Metaphorical Models

The metaphorical models of reading stem from comprehension research conducted over the past three decades. They include bottom-up, top-down, and interactive models.

2.4.1.1 Bottom-up Models

The bottom-up or data-driven models depend on the information given by the text. That information is processed from letter features to letters to words to meaning with little interference from the reader's own background knowledge. Bottom-up models are sequential or linear. The reader processes each word letter-by-letter, each sentence word-by-word and each text sentence-by-sentence. One stage is completed before another is begun (Urquhart and Weir, 1998). The readers' role is just to decode the printed material and they are therefore called a "passive decoder" (Barnett, 1989). Research studies within this model can be focused on small features such as word recognition (Paran, 1996) to larger features such as text organization (Kobayashi, 2002). However, Grabe and Stoller (2002) mentioned that such an extreme view is not entirely accurate.

2.4.1.2 Top-down Models

The top-down models assume that reading is primarily directed by the reader's goals and expectations. The reader is someone who has a set of expectations about the text information. The reader samples enough information from the text to confirm or reject these expectations. Background knowledge is important and inferencing is a prominent feature of these models. Urquhart and Weir (1998) described the reading processes of this model as "reader-driven" processes. However, few reading researchers support strong top-down views for the reason that if a reader must first have expectations about all the information in the text then the reader would learn nothing from the text (Grabe and Stoller, 2002:32).

2.4.1.3. Interactive Models.

Interactive models combine useful elements of both bottom-up and top-down models. However, Grabe and Stoller (2002) commented that word recognition needs to be fast and efficient, but background knowledge is a major contributor to text understanding, as is inferencing and predicting what will come next in the text. They mentioned that the key processing aspects of bottom-up approaches are incompatible with strong top-down controls on reading comprehension and the interactive models turned out to be a self-contradictory model. They further explained that the automatic processing aspects of comprehension, by definition, need to be able to operate without a lot of interference from the moment-to-moment information gained from background knowledge which should be reserved mainly for higher-level processing. They also stated that "modified interactive processes" highlights a number of processes, particularly automatic processes, being carried out primarily in a bottom-up manner with little interference from other processing levels or knowledge resources. Grabe and Stoller (2002: 31) mentioned that though the metaphorical models are useful in the initiation thinking about reading comprehension, "they do not clarify more recent research advances".

2.4.2 Specific Models

In the past twenty years, a number of models have been proposed to be considered as an alternative approach to the metaphorical model. Grabe and Stoller (2002) mentioned four models that have achieved some prominence and have been widely discussed.

2.4.2.1 Psycholinguistic Guessing Model

The Guessing Game Model by Ken Goodman (1996, cited in Grabe and Stoller, 2002) is a strong top-down approach to reading comprehension. The reading process in this view includes (a) hypothesizing, (b) sampling, and (c) confirming information based on background knowledge, expectations about the text, a sampling of surface features of the text and context information from the text. Grabe (2000) mentioned that there is no strong evidence which supports the psycholinguistic guessing model of reading. On the contrary, there is evidence that contradicts the guessing-game model of reading. This model

though well known and popular among applied linguists "is recognized among reading researchers as being fundamentally wrong" (Grabe and Stoller, 2002: 34).

2.4.2.2 Interactive Compensatory Model

The Interactive Compensatory Model (Stanovich, 2000) proposed that (a) reader develops efficient reading processes, (b) less-automatic processes interact regularly, (c) automatic processes operate relatively independently, and (d) reading difficulties lead to increased interaction and compensation, even among processes that would otherwise be more automatic. In other words, in normal reading, what occurs most of the time with fluent readers is a rapid, automatic, context-free word recognition. Readers use context and background knowledge to aid word recognition for better comprehension which is mostly unnecessary for fluent readers. However, poor readers who have deficient word analysis skill might possibly show a greater reliance on contextual factors (Stanovich, 2000: 41). This model is a currently accepted view among psychologists and reading researchers (Grabe and Stoller, 2002; Harrison, 2004).

2.4.2.3 Word Recognition Models

Most Word Recognition Models are based on connectionist theories which focus on how the mind organizes information and learns from the text. These models represent a strong version of the bottom-down process which accounts for word recognition processes under time constraints. Though these models are sources for efficient reading comprehension and for model building, they do not extend the analysis to higher-level comprehension processes (Grabe and Stoller, 2002). Rapid word recognition is important to be a fluent reader. However, the purpose of reading is to gain meaning, not only just to rapidly recognize words (Harrison, 2004).

2.4.2.4 Simple View of Reading Model.

As for the Simple View of Reading Model, Hoover and Gough (1990, cited in Grabe and Stoller, 2002) argue that reading comprehension is composed of a combination of word recognition and general comprehension abilities. This perspective is compatible with word recognition models and the interactive compensatory model. It has generated much discussion

among reading researchers over the past decade (Urquhart and Weir, 1998; Grabe and Stoller, 2002).

In conclusion, the literature on models of reading both metaphorical models and specific models is explored. Interactive compensatory model is a currently accepted view and it proposes that readers depend on their background knowledge when facing reading difficulties. The following two sections explore the differences between the first language readers and the second/foreign language readers and then review the literature about the schema theory.

3. First Language Readers and Second/Foreign Language Readers

Several issues concerning the first language and second/foreign language reading have been studied. On the issue of the Language Threshold Hypothesis, Grabe (2002: 55) mentioned that evidently "second language knowledge plays a much greater role until some general threshold of language knowledge is passed, confirming a general version of the Language Threshold Hypothesis." He also suggested that research found that mental translation as well as the use of dictionaries are useful strategies for second language readers when reading difficult texts. The use of cognates is also an important strategy and the use of glosses provides some benefits to the second language readers.

Another issue is unfamiliar text organization. Grabe (2002) stated that second/foreign language learners need explicit instruction in how texts are structured and how information is organized. He suggested that this is a more general problem of second language exposure where second language readers are seldom exposed to second language texts in the same amount as first language readers.

Though research on second/foreign language reading shows that the factors that influence reading development are quite complex, Grabe and Stoller (2002) have identified three major types of differences between first language and second/foreign language reading contexts and readers. They are linguistic and processing differences, individual and experiential differences, and socio-cultural and institutional differences. Linguistic and processing differences refer to the differences in the amount of lexical, grammatical, and discourse knowledge and the amount of exposure to second language reading, etc. Individual and experiential differences

refer to the differences in the motivations of reading in second language, the differences in the kinds of texts in second language context, etc. Finally, the socio-cultural and institutional differences refer to the differences in socio-cultural backgrounds of the second language readers, the differences in ways of organizing discourse and texts, and the differences in the expectations of second language educational institutions.

In the following section, some of the differences are discussed in more detail in the framework of schema theory.

4. Schema Theory

The term "schemas", or in the Greek plural form "schemata" which is used less widely these days, are mental representations that permit us to store, recognize and remember information (Harrison, 2004). Schema theory proposes that a text does not carry a meaning by itself. The reader brings information, knowledge, emotion, experience, and culture (that is schema) to the printed words. A strong view of schema theory sees them as something influencing the reader's opinion even before a text is read. A weak view refers schema to the organized background knowledge on a topic leading to predictions of discourse (Landry, 2002).

There are two categories of schema: content schema and formal schema (Brown, 2001: 300). Content schema includes all we know about people, the world, culture, and the universe, whereas formal schema consists of our knowledge about discourse structure. However, Landry (2002) mentioned that there is a third category of schema which is abstract schema. Abstract schema refers to hidden factors and thematic considerations. In addition, Barnett (1989) mentioned that content schema or cultural orientation in terms of background knowledge is a factor that influences second/foreign language reading.

The schema theory has been criticized on the basis that it is not a well-defined framework for the mental representation of knowledge. However, it provides an extremely useful notion (Grabe, 1993). This theory describes how prior knowledge is integrated in the reader's memory and used in the higher-level comprehension processes. The implications of schema theory to reading instruction and research have also proven that it is very useful.

The literature that is related to readers that has been explored so far includes communicative competence, the nature of reading ability, the first language readers and the second/foreign language readers, and the schema theory. The role of background knowledge in reading has also been explored. The following section reviews and discusses the literature concerning reading materials or the text.

Text

This section explores the literature related to the text structure and discourse, the text's vocabulary, the text readability, and other factors related to reading texts.

1. Text Structure and Discourse

The literature that has been reviewed so far focuses mainly on the reader. The following section discusses the text itself. In the process of writing a piece of text, writers always choose the best words and the most effective structure. They might return to the text to make changes, to make the text easier for the reader to read and understand, or make the text more accurate or more elegant. The way that the meanings in the text are organized to convey a message is called "discourse" (Nuttall, 1996: 20).

Discourse analysis is the study of how discourse is produced and organized. A closely associated discipline is pragmatics which studies how we use language in particular contexts to achieve particular goals. More specifically, discourse analysis focuses on how the sentences in a text are organized, or how they are related to one another, whereas pragmatics is concerned with how the reader is able to interpret the writer's intention (Nuttall, 1996).

In order to investigate how sentences combine into discourse, the meaning of a single sentence will first be examined. Nuttall (1996) mentioned that there are four kinds of meaning.

1. Conceptual meaning: the meaning that a word can have on its own.
2. Propositional meaning: the meaning that a sentence can have on its own.
3. Contextual meaning: the meaning that a sentence can have only when in context.

4. Pragmatic meaning: the meaning that a sentence has only as part of the interaction between the writer and the reader.

These different kinds of sentences are composed into a piece of language called "text". In order to show that the text is coherent, a number of cohesive devices, such as the words "thus" and "however", are used to make connections and signal relationships. These cohesive devices are also called explicit discourse markers. However, such markers are not necessary if the text is straightforward and the reader can be trusted to identify the meaning without their help (Nuttall, 1996).

The text's coherence depends partly on how sentences are sequenced and on the value that each sentence has. The network of relationship and the way the underlying ideas are organized within a text is called a sentence's "rhetorical structure" (Nuttall, 1996: 26). Some examples of the logical organization of the text are chronological sequence, general statements supported by examples, main ideas and supporting details. Barnett (1989) mentioned that research on formal schemata shows that readers are familiar with standard text structures and they can understand the text if they recognize and follow the structure. Thus, the readers' background knowledge of text structure and discourse cues significantly affect their reading in a second or foreign language (Carrell and Grabe, 2002).

Besides the structure of the text, the type or "genre" of the written text is another important factor in reading comprehension. Genre is the set of governing rules and writing conventions applied depending on each type of text. Barnett (1989: 119) mentioned that "many text genres are well-established forms with predictable structures and, sometimes, predictable content". Brown and Yule (1983) suggested that genres may differ in their paragraph structure, thematic sequence, the stereotypical ordering of events in time, and the distribution of sentence types and pronominal forms, etc. Brown (2004) mentioned that readers need to apply certain schemata to assist them in extracting the appropriate meaning from different genres. He gave the example that in reading a recipe, the reader knows that it provides a list of ingredients and a sequential order of how to cook them.

Another subject matter of the reading text is the "topic" which tells the readers the subject of the text. It is clear that favorable topics draw readers' interests

and familiar topics facilitate readers to read. As mentioned in the previous section dealing with schema theory, it is clear that content schema is certainly related to the topic of the reading text (Landry, 2002, Carrell and Grabe, 2002).

In this section, various studies are explored and discussed in the areas of the text structure and discourse which also include the areas of text cohesion, the genre, and the topic of the text. The following section will explore a smaller unit of the text which is the vocabulary.

2. Vocabulary

Researchers and reading experts agreed that reading comprehension is dependent on word knowledge (Alderson, 2000a; Nation, 2001; Qian and Schedl, 2004, Harrison, 2004). Grabe (2002: 57) mentioned that "vocabulary knowledge is at the heart of fluent reading abilities. A large recognition vocabulary is essential".

It is suggested that vocabulary knowledge should at least comprise two dimensions, which are vocabulary breadth and vocabulary depth (Qian and Schedl, 2004). Vocabulary breadth or size refers to the number of words that a reader has at least some superficial knowledge of their meanings. Vocabulary depth or quality refers to how well a reader knows a word.

In reading assessment, vocabulary knowledge is considered an important factor (Qian and Schedl, 2004). An example is in the construction of the TOEFL. Vocabulary has a major role in determining tasks and item difficulty (Enright et al., 2000).

In this study, word knowledge is considered as an important part of assessment of student reading comprehension. A large proportion of the designed test items is constructed to measure the test-takers' skill in identifying the meaning of words.

3. Text Readability

In the process of material preparation either for instructional or testing purposes, teachers or testers have to estimate the difficulty of the text in order to select the readable text for their students. Studies have been carried out to identify the factors that might involve the readability of the text and formulas have been

developed to estimate the difficulty of the text. Many of the formulas attempt to measure the semantic complexity of a text by assessing the number or percentage of difficult words in that text. Difficult words or hard words are usually defined as unfamiliar words or words with several syllables (Irwin, 1991:155). Some readability indices, therefore, use word length as the indicator of the difficulty of the text. One of the indexes of word length is the "FOG" index (Alderson, 2000a: 71). It indicates the difficulty of the text by counting the number of words in text containing three or more syllables.

Other readability indices are related to both vocabulary complexity and sentence length. One of the most popular and easy to use formulas of this kind is the Fry Readability Formula. It determines readability by randomly select samples of 100 words, counting the number of sentences and syllables, and averaging these counts. It then locates the average number of sentences on the vertical axis and the average number of syllables on the horizontal axis (see Figure 1). The point on the graph where these two lines intersect indicates the approximate readability, expressed as a grade-level equivalency (O'Donnell and Wood, 2004: 155).

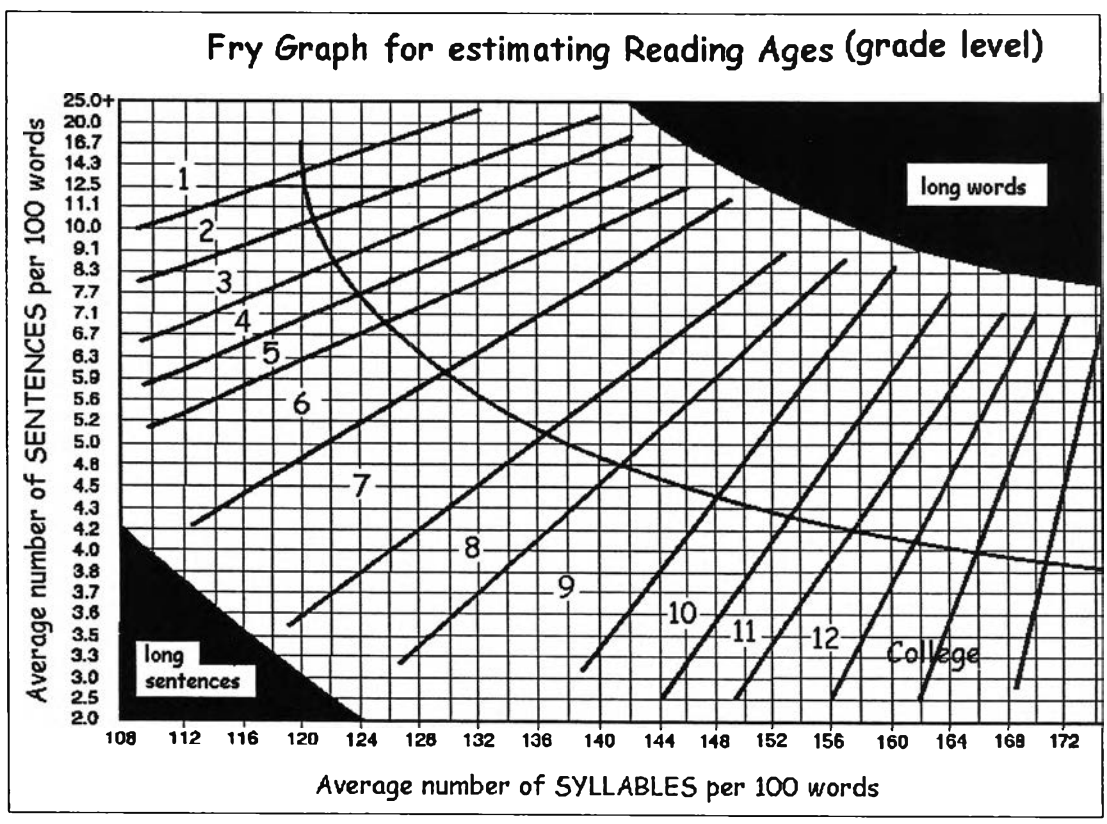

Figure 1. The Fry Graph

Another frequently used readability formula is Flesch Reading Ease formula. It was developed in 1948 and is still in use today (Alderson, 2000a: 71). The Flesch Reading Ease score can be calculated by:

$$RE = 206.835 - (0.846 \times NSYLL) - (1.015 \times W/S)$$

where NSYLL is the average number of syllables per 100 words and W/S is the average number of words per sentence.

The Flesch Reading Ease indices can be calculated by the Microsoft Word for Windows. After Microsoft Word completes a grammar check on a text, the readability statistics are then automatically displayed.

Schnelbach and Wyatt (2005) mentioned that the Flesch Reading Ease score can be matched with the difficulty and the grade level of the students. The Flesch Reading Ease score of 0-29 is a "very difficult" text which is suitable for post graduate students. A score of 30-49 is described as a "difficult" text suitable for college students and a score of 50-59 is "fairly difficult" which is regarded as suitable for high school students. The highest score is a score of 90-100 which means that the text is "very easy" and is suitable for 4^{th} to 5^{th} grade students. Thus, the lower the Flesch index, the more difficult the passage is deemed to be.

The readability index provides only a rough guide because it cannot take account of the concept density of the material (O'Donnell and Wood, 2004: 155), the structural and rhetorical feature of the text (Clapham, 1996), and a reader's knowledge and interest (Clapham, 1996, Nuttall, 1996: 176). We are not yet able to calculate a score or index for these essential factors. They can only be determined by reading experts, teachers, testers, who are familiar with the text and with the backgrounds of the students or test-takers who are expected to read it (Clapham, 1996).

4. Other Factors Related to the Text

Other factors which potentially affect reading comprehension include typographical features, verbal and non-verbal information, and the medium of text presentation (Alderson, 2000a).

The typographical features include the layout of the print on the page, the type of the font, the size of the font, the density of word per line, etc. Alderson

(2000a) mentioned that although the effect of these factors on second or foreign language reading is not apparent, it is advised that the texts are suitably presented. He added that it is obviously undesirable for readers to be penalized by poor or untypical text layout or reproduction.

Though most reading texts tend to contain only verbal information, non-verbal or graphic information is increasingly used in reading texts. Non-verbal information includes pictures, tables, graphs, diagrams, charts, etc. Some of them are better in attracting readers' attention and some of them last longer in the readers' mind.

The last factor to be considered is the medium of text presentation which includes textbooks, overhead projector slides, television screens, LCD projectors, or computer screens. Alderson (2000a) mentioned that information is increasingly available on computer screens, especially with the development of the Internet and the World Wide Web, and the use of computer-based self-instruction.

In this part we have considered the literature in the areas of the text and discourse, the text's vocabulary, the text's readability, and other factors related to reading texts. The information from literature reviews are employed in the construction of the reading comprehension computer-based test which is one of the tools for collecting data about test-takers' reading comprehension CBT ability. This data will be used in the correlational analysis which is intended to yield answers to the research questions and at the same time test the research hypotheses.

The following section reviews the literature on language testing.

Language Testing

Brown (2004: 3) defined a test as "a method of measuring a person's ability, knowledge, or performance in a given domain". He mentioned that according to the definition a test is first a **method**. This method must be explicit and structured. Examples are multiple-choice questions with prescribed correct answers, a writing prompt with a scoring rubric, etc. Secondly, a test **measures** either general or specific competencies or abilities. It provides one or more forms of reporting measurement or results. Thirdly, a test measures an **individual's** ability, knowledge, or performance. Test writers need to know who the test-takers are and write the test to match their

abilities, experience, background, etc. Fourthly, a test measures **performance**. The results imply the test-takers' ability or competence. It can be designed to measure the test-takers' ability to perform a language e.g. the ability to speak, read, write, or listen. It can also be designed to test the knowledge about the language of the test-takers e.g. the ability to define vocabulary items, reciting grammatical rules, etc. Brown (2004) explained that performance-based tests sample the test-takers' actual use of the language and make inferences about their competence from those samples. Therefore, a typical test of reading comprehension consists of short reading passages followed by a number of comprehension questions. The passages in the test are a small sample of a second/foreign language of the reader's total reading behavior. However, testers can infer about test-takers' general reading ability from the results of the test. Finally, a test measures a given **domain**. While a proficiency test's domain may measure general competence in all skills of a language, other tests may be very specific. For example, a vocabulary test may focus on only a set of words in a particular area.

The trends and practices of language testing in the past have followed the shifting views of the language teaching methodology. In the 1950s, language teaching was dominated by the view of behaviorism where contrastive analysis received special attention. In that era, testing focused on specific language elements such as phonological, grammatical, and lexical contrasts between two languages. In the 1970s and the 1980s, language teaching was in the era of communicative theories where language testing held a more integrative view. At the present time test writers seek to produce a more authentic, valid instrument that simulates real-world interactions (Brown, 2004).

The reading frameworks underlying the tests or definitional constructs are put into the explicit and operational one through the "test specifications" (Alderson, 2000a). Theoretical frameworks that have been used in developing the test specifications included the Munby model (1978), and the Bachman model (Bachman, 1990), etc.

Furthermore, a model of test usefulness was proposed by Bachman and Palmer in 1996 as the basis for quality control in the test developing process. The test usefulness includes six test qualities which are reliability, construct validity,

authenticity, interactiveness, impact, and practicality. They also discussed in detail two sets of characteristics, the language use tasks or test tasks and the language users or test-takers. These are considered in the following section.

The Test Characteristics

Bachman and Palmer (1996) believed that there are two major factors or characteristics that affect both language use and test performance. Those characteristics are test task characteristics and test-takers' characteristics.

1. Test Task Characteristics

Bachman's (1990) framework of "Method Facet" has been questioned about its practicality. Bachman and Palmer (1996) based on Bachman's (1990) framework developed an operational concept of "Task Characteristics." They extensively discussed and gave examples of how this framework can be applied to practical problems in designing and developing language tests.

Their framework of task characteristics involves five areas which are

1. characteristics of the setting,
2. characteristics of the test rubrics,
3. characteristics of the input,
4. characteristics of the expected response, and
5. relationship between input and response.

Bachman and Palmer (1996) mentioned that the characteristics of the setting are concerned with the physical environment where the test takes place. They include the physical setting, participants, and time of the tasks. The physical characteristics are the location, noise, temperature, humidity, seating condition, lighting and degree of familiarity of materials and equipment to the test-takers. They gave an example that in an office setting familiarity with materials and equipment, such as a computer system, clearly affect an individual's use of a language. The participants in the setting are test-takers and test administrators. Finally, the time of the test task is when the test is administered. When the test takes place, are the test-takers fresh or tired?

The characteristics of the test rubrics include the structure of the test, the instructions, the duration of the test, and the scoring method. The instructions must be explicit and inform test-takers about the procedures for taking the test, how it will be scored, and how the results will be used. The structure of the test concerns how the parts of the test are organized and presented to the test-takers. Questions include how many parts and how many items should be included, a single or multiple tasks, and which parts or tasks come first?, etc. The duration of the test is the amount of time allotted for each task and for the entire test. Is it a speed or power test? The scoring method includes how the responses are determined, by objective scoring key, or one rater, or multiple raters, etc.

The characteristics of the input are the format and the language of the input. The format concerns how the tests are presented. They should include the channel, form, language, type of input, degree of speededness, and the vehicle. The language of input includes the areas of language knowledge and topical knowledge.

Similarly, the characteristics of the expected response are the responding format and the language of expected response. The format includes the channel, form, language, length, type, and degree of speededness. The language of the expected response has the same characteristics as that of the language of input.

Finally, the relationship between input and response are concerned with the reactivity, scope, and directness of the relationship. Reactivity is the degree that the response affects the subsequent response. The scope of relationship is the amount of input that test-takers need to process in order to respond. Reading for main ideas is an example of a test task that requires the reader to read the whole passage to understand the gist of the text. A multiple-choice grammar item is an example of a micro skill test task. The directness of relationship is the degree to which the test-takers rely on information in the context/topical knowledge or just on the supplied input. A speaking test that requires test-takers to give their opinion about current social issues is an example of indirect relationship whereas a speaking test that requires test-takers to describe the content in a picture is an example of the direct relationship.

2. Test-Takers' Characteristics

Bachman and Palmer (1996) mentioned that a large number of individual characteristics affect the language use by the test-takers. Many of those characteristics such as fatigue or sudden mood shifts are unpredictable and test writers almost cannot design their tests to accommodate these characteristics. Nevertheless, Bachman and Palmer (1996: 64) proposed four sets of individual characteristics whose effects on language test performance are better understood and suggested that in designing, developing, and using language tests, test writers should give attention to these four sets of characteristics.

1. Personal characteristics. Bachman and Palmer (1996) mentioned that it is clear that to give a complete list of personal characteristics is impossible. However, they provided some test-takers' characteristics as a starting point. The suggested characteristics are age, sex, nationality, residence status, native language, level and type of general education, and the type and amount of preparation or prior experience with a given test.

2. Topical knowledge. This refers to knowledge schemata or real-world knowledge. For example, for a Thai student, a reading passage that includes an American cultural specific might be more difficult to read than a reading passage that includes a Thai cultural specific.

3. Affective schemata. This refers to the affective or emotional responses that are associated with topical knowledge. It can either facilitate or limit the performance of the test-takers in a given context. For example, a controversial topic, such as abortion, may stimulate some individuals to perform at a high level.

4. Language ability. Bachman and Palmer (1996) adopted Bachman's (1990) model of language ability which involves two components: language knowledge and strategic competence. The combination of these two provides test-takers with the ability to create, interpret, and respond to the test tasks.

Bachman and Palmer (1996: 39) commented that the test task and the test-takers have to be interactive, given that interactivness is an important test quality. They described interactiveness as a function of the extent and type of involvement of the test-takers' charactertistics in accomplisthing a test task.

Reading Tests

The following section focuses on the assessment of reading comprehension and the questions for reading tests.

1. Assessing Reading Comprehension

In any reading test specification, a list of reading skills which will be tested is usually given. However, to write test items which follow such requirements is difficult because most reading items test a variety of skills, and different readers use different processes or skills when they read a text (Clapham, 1996).

Furthermore, there is no concensus among reading experts about the taxonomy of reading skills. Examples of the most widely used taxonomies of supposed reading skills include Davis's (1968, cited in Alderson, 2000a) defined eight skills and Munby's (1978, cited in Alderson, 2000a) taxonomy of reading "microskills". The taxonomy of reading skills has been influential in the design of the language tests (Alderson, 2000a).

Nevertheless, there are also many criticisms of the use of taxonomies. Such criticisms them include the lack of empirical data to support taxonomies and the lack of discrete definitions. Reading involves several overlapping skills, which are used in conjunction with each other as necessary and it is almost impossible to isolate specific skills that are used in response to specific test items. Furthermore, the analysis of test performance does not support such separation of skills (Alderson, 2000a: 11).

Subsequently, many researchers focused on the process approaches to reading. However, there is "a growing realization that processes of reading are more complex than originally assumed, and the inevitable pendulum swing in research and teaching fashions, have led to revived interest in the product of reading" and "a skills approach to defining reading remains popular and influential and cannot be ignored in a treatment of the nature of reading" (Alderson, 2000a: 5, 11).

In the process of writing test items, a list of skills is required by the test writers. This list is used to identify the ability that the test-takers are to be able to demonstrate in order to show that they understand the text and are able to reach the correct answer. Though taxonomies are heavily criticized, they potentially provide a

very powerful framework for test construction and are likely to be continually used by test writers (Alderson, 2000a).

The International English Language Testing System (IELTS) is a standardized test which is administered worldwide. The subskills included in the IELTS Bands 5, 6, and 7 that are in the reading section are listed below.

1. Identifying structure, content, sequence of events and procedures
2. Following instructions
3. Finding the main ideas which the writer has attempted to make salient
4. Identifying the underlying theme or concept
5. Identifying ideas in the text and the relationships between them
6. Identifying, distinguishing and comparing facts, evidence, opinions, implications, definitions and hypotheses
7. Evaluating and challenging evidence
8. Formulating a hypothesis from an underlying theme, concept and evidence
9. Reaching a conclusion by relating supporting evidence to the main idea
10. Drawing logical inferences

Though each skill is listed as if they are independent concepts, this does not mean that each of the abilities listed above can be tested separately or in isolation (Clapham, 1996).

In the context of Thai students studying English as a foreign language, Phakiti (2003) gave a list of objectives of a Thai university fundamental English course. In that list the following reading comprehension skills were emphasized for the purpose of assessing students.

1. Scanning and skimming text for general and specific information
2. Finding answers explicitly or implicitly to questions
3. Recalling word meaning
4. Skimming the text to evaluate information
5. Guessing meanings of unknown words from context clues
6. Identifying phrases or word equivalence

7. Predicting topics of passages and the content of a passage from an introductory paragraph
8. Recognizing the abbreviations of words
9. Making decisions for appropriate information
10. Discriminating between more or less important ideas
11. Discriminating facts from opinions
12. Analyzing reference words
13. Drawing inferences from the context
14. Identifying the title of the text and the appropriate heading
15. Summarizing the content of the text
16. Finding the main ideas of the text

Again, although each item in this list appears to be an independent concept, it can hardly be tested separately or in isolation. Furthermore, this list cannot claim to be inclusive of all reading comprehension subskills.

Though there are some problems with the use of taxonomy as discussed, it is still influential in writing language tests (Alderson, 2000a). Therefore, some of the commonly used reading comprehension subskills previously listed have been selected to be included in the reading test of this study. The selected subskills are identifying word meaning, identifying explicit information/opinions/definitions/facts, analyzing reference words, drawing inferences from the content, finding the main idea and topic, and understanding the writer's intention and attitude.

2. Questions for Reading Tests

The questions used to ask in the reading test can be classified into six types (Nuttall, 1996).

1. Questions of literal comprehension
2. Questions involving reorganization or reinterpretation
3. Questions of inference
4. Questions of evaluation
5. Questions of personal response
6. Questions concerned with how writers say what they mean

The answers to the first type of questions are explicitly expressed in the text. Type 2 questions require the students either to reinterpret literal information or to obtain it from various parts of the text and put it in their own way. Type 3 questions oblige the students to consider what is implied in the text. Similar to type 2 questions, they may require readers to put together scattered information, but type 3 questions demand more sophisticated inferencing skills. Type 4 questions require the readers to make a judgment about the text in terms of what the intention of the writer is and how effective the narrative power is. Type 5 questions ask the readers to record their reactions to the text. Finally, type 6 questions are concerned with how the writer says and what he or she means (Nuttall, 1996: 188-189).

To write questions for the reading test, the test writer should take the following items into consideration (Fillmore and Kay, 1983, cited in Urquhart and Weir, 1998).

- Questions should not contain harder vocabulary than the text.
- Questions should have only one unequivocal answer.
- If the candidates understand the text, they should be able to answer the question.
- Rejection of alternatives on grammatical grounds should not be allowed.
- Skills not related to reading, e.g. mathematics, should not be tested.
- Incidental insignificant information should not be tested.
- Questions that require stylistic or other ambiguous judgments should be avoided.

These guidelines and the suggestions together with comments from content experts have been taken into consideration in the process of writing and revising the test items in this study.

Computerized Tests

This section discusses the advantages and disadvantages of computerized tests (any tests that use the computer as the mode of delivery) and the comparability of the computerized tests and paper-and-pencil tests. The differences between two subsets

of the computerized tests: the computer-based tests and the computer-adaptive tests are also addressed. Finally, the issues related to computerized reading comprehension tests is explored.

1. The Advantanges and Disadvantages of Computerized Tests

The advantages and disadvantages of the computerized tests have been reported in many studies. Alderson (2000b) explained that the advantages of computerized testing include the flexibility of testing dates and locations, the immediate availability of the results, the ease of item authoring and reviewing, and the enhancement of test security. Cohen (1994) listed the advantages of computerized assessment as follows:

1. Individual testing time may be reduced.
2. Frustration and fatigue are minimized.
3. Boredom is reduced.
4. Test scores and diagnostic feedback are provided immediately.
5. Respondents can be given a second chance to get an item right.
6. Test security may be enhanced if respondents receive different items or the same items in a different sequence.
7. Accurate and consistent evaluation of results is provided.
8. Record-keeping functions are improved.
9. Teachers are provided with diagnostic information.
10. There is rapid access to banks of test items.
11. There is a potential for rapid editing of items.
12. Information is readily available for research purposes.

There are also some disadvantages in applying productive skills of speaking and writing to the computerized tests. Test-takers should be aware of those disadvantages and need to be familiar with the computer and their limitations (Alderson, 2000b). Some disadvantages of computerized tests are listed by Brown (1997) which include:

1. Computer equipment may not always be universally available, or in working order.

2. The limitation of the amount of material can be presented on a computer screen.
3. There is the limitation of graphic capabilities of computers.
4. The changing mode of presentation may affect students' performance.
5. Different computer familiarity may affect the performance.
6. Computer anxiety also has the potential to affect the test results.

2. The Comparability of the Computerized and Paper-and-Pencil Tests

Besides the advantages and disadvantages of the new mode of testing, numerous amounts of studies focus their investigations on the comparability of the computerized and paper-and-pencil tests.

Mead and Drasgow (1993), for example, examined the effects of the paper-and-pencil versus computerized modes of test administration. They conducted a correlational study on timed power test and speeded test of cognitive abilities for populations of young adults and adults. The cross-mode correlation was estimated to be .72 for speeded tests and .97 for timed power tests. The results of their study provide strong support that a substantial medium effect was found for speeded tests but there was no mode effect for carefully constructed power tests.

Nevertheless, Neuman and Baydoun (1998) examined the cross-mode equivalence of paper-and-pencil and computer-based clerical tests among undergraduate students of two universities by administering 10 timed clerical tests. They found no significantly differential validity across different test formats at the alpha level of .0005 and the paper-and-pencil and the computer-based test performances are consistently and highly cross-mode correlated. The structural equation modeling suggested that the constructs being measured in the paper-and-pencil and the computer-based test were equivalent.

3. Computer-Adaptive Tests VS Computer-Based Tests

The computerized test is a term that refers to any test that is delivered via the computer mode. The simplest form of computerized tests is the Computer-Based Test (CBT) which is a linear computer-delivered test. The CBT has been

criticized on the ground that it is not tailored for the specific ability of an individual student. It delivers the same set of questions to all participants of the test session, regardless of their ability. Students with high performances are presented with one or more questions that are below their level of ability, while students with low performances are presented with one or more questions that are above their level of ability. Low proficient students might experience frustration or anxiety when they are presented with difficult questions, while high proficient students might feel that the test is less than challenging if the questions are too easy.

The Computer-Adaptive Test (CAT) which is sometimes called a "tailored test" (Sukamolson, 2003), is based on the Item Response Theory (IRT). The CAT offers a set of questions that is appropriate to each student's level of ability. It selects the questions based on an individual student's performance during the test session. In general, the CAT starts with a random question of average difficulty for the test population. After the students have responded, "the next best item" is presented (Young et al., 1996). If the student answers the question correctly, the estimate of his or her ability is increased; thus, a more challenging question is selected to be the next presented item. However, if the student gives a wrong answer, the estimated ability of the student is decreased, and an easier question is presented as the next question. By using this method, the test length (test items) and the testing time can be reduced, and thus also the testing fatigue (Sukamolson, 2003; Lilley, Barker, and Britton, 2004). In addition, the CAT also enhances test security because different individuals take different tests (Young et al. 1996, Alderson, 2000b).

Although the CAT seems to be a better choice in the delivery of tests there are some limitations. In the development and evaluation of a computer-adaptive test in higher education to estimate the level of proficiency in English for students whose first language is not English, Lilley et al. (2004) compared the performance between CAT and CBT. They found that the CAT was at least as useful as CBT. They provided evidence that students are not disadvantaged by the CAT approach. Nevertheless, they mentioned that the CAT is more difficult to construct than the CBT. The CAT requires an adaptive algorithm and a larger calibrated question bank (Alderson, 2000b, Lilley et al., 2004). Brown (2004) mentioned another limitation in the CAT process which is that the test-takers see only one question at a time which

the computer scores before selecting the next question. Thus, test-takers cannot skip any question. In addition, Alderson (2000a: 100) and Brown (2004) stated that after the test-takers enter and confirm their answers, they cannot go back to the questions or to any earlier part of the test. Thus, they cannot review and revise their answers.

4. The Computerized Reading Comprehension Tests

There are some difficulties when applying CAT to the traditional paper-and-pencil tests. Firstly, the students have to answer the questions delivered to them based on the level of difficulty of the items regardless of the passages on which the items are based (Wainer and Kiely, 1987, cited in Young et al., 1996). For example, the first two items may be based on the first passage while the third item might ask a question based on the second passage and then the fourth and the fifth items may switch back to ask about the first passage again. Wainer and Kiely (1987: 188, cited in Young et al., 1996: 37) proposed a "testlet-based CAT" where bundles of items that are treated as a unit replace item banks. They explained that the testlet is a group of items related to a single content area and contains a fixed number of predetermined paths for test-takers to follow. Students are branched from one testlet to another based on their estimated abilities when they have completed all items in one testlet. Thus, students do not have to switch backwards and forthwards between passages like they do when reading the passage in the item-based CAT.

Another issue with the CAT is that it assumes that each item is independent of its context and this is not a reasonable assumption (Johnston, 1984, cited in Young et al. 1996: 170). He explained that the order in which questions are presented tends to influence the manner in which individuals respond to them, in inducing cognitive response sets.

Furthermore, Young et al. (1996) mentioned that it is accepted in testing practice to deliver easy items at the beginning of a test because it is believed that students are more comfortable and feel less anxious if their initial attempts are successful and so are more likely to perform to their true competence.

Test-takers' variables

As the use of computer-based tests increases, especially in high-stakes educational assessments the concern about the equivalence between the computerized tests and the paper-and-pencil tests increases. Though many studies have been conducted to investigate the equivalence of the two modes of test, McDonald (2002) mentioned that the studies of score equivalence have largely ignored individual differences. He reviewed research studies and related literature and concluded that there are three computer related individual differences which are significantly related to the testing mode. They are computer familiarity, computer anxiety, and computer attitudes. The studies concerning these computer-related variables are presented below.

1. Computer Familiarity

Quite a number of studies have been conducted on "computer familiarity," which is also often referred to as "computer experience." There is little agreement concerning the definition of the term "computer familiarity" among researchers (Smith et al., 1999). Most of them came up with different dimensions and focuses.

Some studies focused on computer use as the only indicator of computer familiarity. They referred to computer familiarity as the years or months of computer use, or hours per week, etc. that a test-taker had experienced (Weil and Rosen, 1995; Hong and Koh, 2002; van Braak, 2004). Some researchers included computer courses as part of computer familiarity (Bradley and Russell, 1997; Rosen and Weil, 1995a). Karsten and Roth (1998) associated computer familiarity with the total number of years of computer experience, current average hours per week of computer use, and also the number of prior computer courses completed by a test-taker.

Nash and Moroz (1997) on the other hand, presented a quantification of experience in terms of frequency and intensity, rather than one based on a summary of time-related experience. Alternatively other studies have focused on the ability of the computer users in working with the software and hardware of the computers (Potosky and Bobko, 1998, Russell, 1999).

Other researchers have regarded the availability of the computer as an indicator of computer experience. They found that computer availability at home is

one of the significant factors that contributes the most to the acquisition of computer experience (Looker and Thiessen, 2003; Attenwell and Battle, 1999).

Jones and Clark (1995, cited in Garland and Noyes, 2004) defined computer experience in three dimensions. They are amount of computer use, opportunities to use computers, and the diversity of computer experience. The opportunity to use computers includes issues such as whether a person can obtain access to a computer at home or university, the extent to which they use computers, and whether they have ever participated a course that required the use of a computer (Brosnan and Lee, 1998).

In reviewing 16 studies which are concerned with the concept of computer familiarity, Taylor et al. (1999) mentioned that the term was variously defined and included different dimensions which are experience, frequency of use, type of use, number of courses involving computers, owning a computer, access to computers, attitudes toward computers and related technologies.

Other researchers measured computer familiarity in terms of "skill level". For example, by asking respondents to rate their skills on a range of computer tasks such as typing, programming, word processing, spreadsheets, using the Internet, using e-mail, and finding information online (Taylor et al., 1999; Bozionelos, 2001b; Schumacher and Morahan-Martin, 2001; Stricker, Wilder, and Rock, 2004). Hasan (2003) called an individual's perceptions about his or her own ability to use a computer to perform a computing task successfully as "computer self-efficacy". Hasan (2003) investigated the influences of computer experience on computer self-efficacy and found that the different types of computer experience or knowledge has different magnitude on computer self-efficacy beliefs.

Another way of assessing computer skills is to ask about perceived computer knowledge. For example, Levine and Donitsa-Schmidt (1998) asked respondents if they understand the computer terminology. While Potosky and Bobko's Computer Understanding and Experience scale (CUE) covered the same basis and also included questions on computer literacy and questions like "how good the respondents feel they are at using computers?" (Potosky and Bobko, 1998). They defined computer experience as "the degree to which a person understands how to use a computer. That is, an experienced computer user understands enough about computers in order to use

them, more or less independent of specific software packages, reasons for use, and computer hardware features" (Potosky and Bobko, 1998: 338).

In examining the relationship between computer familiarity and performance on computer-based language test, Taylor et al. (1999) included the number of time examinees take computer-based tests into the construct of computer familiarity. This reflects in their rating scale questionnaire which includes a question "How many examination/tests have you taken on a computer?"

More recently, Beckers and Schmidt (2003) measured computer experience in their two studies in terms of breadth of experience, the number of hours spent working with computers, skill level, the nature of the first computer experience, and the occurrence of computer anxiety.

Therefore, computer familiarity can be seen as the sum of all computer-related events. These events are (1) the year/month of computer use (2) the frequency of use, e.g. per hour, day, week, month (3) the hardware and software used such as a personal computer or a personal digital assistant and applications such as word processing, databases, programming, e-mailing, downloading software from the Internet (4) the number of computer related course taken (5) computer access points such as at home, at university, or at Internet café (6) owning of a computer (7) Internet experience (8) computer self-efficacy (9) computer anxiety (10) number of time taken computer-based tests.

Some dimensions of the reviewed computer familiarity which are relevant to the context of this study are selected. Based on the previously discussed studies, "Computer familiarity" in this study refers to (1) the frequency of use which includes place of access (at home, university, and Internet café), (2) the purpose of use (for educational, recreational purposes, or Internet access), (3) the ability and skills which include perceived ability (to use computer software and hardware), and (4) the computer related skills (computer-based tests and years of computer use).

2. Computer Anxiety

Over several decades "Computer phobia", which is a widespread phenomenon in the student population, has been increasingly recognized (Selwyn, 2000). In 1995, Weil and Rosen studied computer anxiety in university students from 23 countries and

found that the percentage of students who could be classified as computer phobic (ranging from low to high), was approximately 25%. A subsequent meta-analysis by Chua, Chen, and Wong (1999) indicated that there has been no reduction in this problem over recent years.

"Computer anxiety" or "computer phobia" is "a fear for computer when using the computer or when considering the possibility of computer use" (Chua et al., 1999). The term relates to emotional fear, apprehension and the phobia of individuals who use computers or think about using computers. It is characterized as an affective response and an emotional fear of potential negative outcomes such as damaging the equipment or looking foolish.

The terms computer anxiety, computerphobia, and computer aversion have been used to describe "the negative reactions of individuals who experience bad feeling and agitations in the presence of, interacting with, or thinking about computers" (Gaudron and Vignoli, 2002). Studies have confirmed that high computer anxiety will reduce a person's effectiveness when using a computer. As such, the performance of computer tasks by students with high computer anxiety might be poorer than those students that have little or no computer anxiety (Heinssen, Glass, and Knight, 1987; King, Bond, and Blandford, 2002; Barbeite and Weiss, 2004,).

Psychologists have classified general anxiety into two areas which are "trait" and "state" anxiety. Trait anxiety refers to "a general readiness to react with anxiety in many situations" and state anxiety refers to "anxiety actually experienced in a particular situation" (Biggs and Moore, 1993, cited in King et al., 2002: 11). Computer anxiety and test anxiety are examples of state anxiety.

State anxiety can be changed (Chua et al., 1999; Rosen, Scars, and Weil, 1993; Heinssen et al., 1987) while trait anxiety is more persistent and may become a part of the individual's psychological status (Yaghi and Abu-Saba, 1998). According to the interaction model of personality, the trait anxiety is multidimensional. The trait anxiety and the congruent situation factors interact to determine appraisal of situational threat, resulting in state anxiety. Individuals, thus, differ in anxiety proneness with respect to certain types of situations (Gaudron and Vignoli, 2002).

Anxiety can generate a range of emotional responses which can be identified by two methods. The first method involves detecting actual physiological changes in

the subject while the second method requires the subject to complete self-report tests. The second method is more practical in educational environments and generally, it uses a Likert scale format and records a person's perceived feelings, attitudes and reactions as opposed to how their bodies are physically responding (King et al., 2002).

Over the past three decades, many studies have been carried out in the area of computer anxiety which is a psychological phenomenon (Beckers and Schmidt, 2001). Different aspects of computer anxiety have been explored. Some studies have focused on the development of measures and the relationship to other personality or demographic variables. Others examined the relationships with other factors, such as test anxiety, maths anxiety, or computer experience (Namlu, 2003).

A variety of self-report measures have been developed to assess the level of computer anxiety. In 1970, C. D. Spielberger (cited in Truell and Meggison, 2003) developed a State-Trait Anxiety Index to measure both a person's state and trait anxiety level. Based on the interaction model of anxiety, Gaudron and Vignoli (2002) developed a measure of computer anxiety. The self-report measures are separated between the state and trait anxieties. After the validation process, they concluded that the data supported the reliability and the validity of the constructed scale.

In 1984, Loyd and Gressard developed a ten item computer anxiety scale as a sub scale of Computer Attitudes Scale (CAS). Participants are asked to rate a four Likert scale if they strongly agree, slightly agree, slightly disagree, or strongly disagree with the ten statements.

In 1987, Heinssen et al. developed a 19-item self-rating assessment of an individual's level of computer anxiety called a Computer Anxiety Rating Scale (CARS) to measure the response to interaction or anticipated interaction with computers. The scale measures resistance to and avoidance of computer technology as a function of fear and apprehension, intimidation, hostility, and concerns that individual will be embarrassed, look stupid, or even damage the equipment. The statements in the scale reflect perspectives mentioned above such as the fear (e.g. I feel apprehensive about using computer.), the attitude (e.g. I dislike working with machines that are smarter than I am.), the perceived ability (e.g. I am confident that I can learn computer skills.), etc.

Other studies conducted by Rosen and Weil's (1995b) and Dyck, Gee and Smither's (1998) focused on the circumstances under which computer anxiety emerges (i.e. Does anxiety only appear when actually dealing with a computer, or does it already emerge while thinking about using it or seeing others use it?).

Rosen and Weil's (1995b) Computer Anxiety Rating Scale is used extensively by researchers and educators to ascertain the level of computerphobia among students. However, Gordon et al. (2003) pointed out that the three-factor model of the CARS is regarded as a poor explanation of the sample data consisting of 661 undergraduate students in five universities of the United Kingdom.

In Australia, Bradley and Russell (1997) studied the computer anxiety among Australian school teachers and reported three types of computer anxiety: damage anxiety – the fear that they would damage the computer; task anxiety – the fear of performing computer tasks; and social anxiety – the fear that they would embarrass themselves.

Beckers and Schmidt (2001) reviewed seven studies that employed factor analysis to explore underlying dimensions. They concluded that computer anxiety is composed of at least the following elements: (1) low confidence in one's own ability to use computers; (2) negatively affective responses to them; (3) becoming aroused while using a computer or thinking about it; and (4) negative beliefs about the role of the computer in our lives. They tested the six-factor model of computer anxiety in two samples of university students. The dimensions involved were: computer literacy, self-efficacy, physical arousal caused by computers, affective feelings about them, beliefs about the beneficial effects of computers, and beliefs about their dehumanizing aspects. They found that computer literacy has a strong directional influence on both physical arousal and affects. Beliefs about computers were shown to be dependent on affects and physical arousal. From the findings they suggested that training programs that enhance self-efficacy and computer literacy may in principle reduce computer anxiety. Levine and Donitsa-Schmidt (1998) even argued that computer self-confidence and computer anxiety are essentially the same thing.

In Chou's (2001) experimental study, which included the investigation of the effects of computer anxiety on self-efficacy, the concept of computer anxiety had a significant effect on computer self-efficacy.

In summary, computer anxiety can be seen as different dimensions which include (1) the fear of the computer (thinking about or working on it), (2) resistance to the computer, (3) avoidance of the computer, (4) fear of damage the computer, (5) fear of loss of face (social anxiety), (6) dehumanizing aspects of the computer, and (7) computer self-efficacy.

Some dimensions from the above are selected to be components in this study. Some are excluded because they are regarded as being related closer to the computer attitud factor than the computer anxiety factor. In this study, the term "computer anxiety" refers to the fear experience when interacting with a computer or anticipating an interaction and the fear that relates to the lack of confidence to use a computer and to learn something new about the computer.

3. Computer Attitudes.

Noyes and Garland (2005) mentioned that although a large number of studies have been conducted on measuring attitudes towards computers, there is little agreement on the definition of the term. Kay (1993) noted that "computer attitudes" has been defined in over 14 different ways.

The instruments designed to measure computer attitudes dated back to Lee's (1970, cited in Shaft, Sharfman, and Wu, 2004: 55) study of social attitudes towards "electronic brain machines". Since then, instruments for measuring attitudes towards computers have been continuously constructed by researchers and educators. Shaft et al. (2004) overviewed various constructed instruments to assess computer attitudes and presented 31 instruments in chronological order. Some of them are selected, reviewed and discussed in this study.

In 1984 Loyd and Gressard developed a Computer Attitude Scale (CAS) to gather information concerning attitudes towards learning about and working with computers. It comprised three dimensions namely computer anxiety, confidence, and liking. In 1985 Loyd and Loyd constructed a 4-dimension CAS. The fourth dimension is computer usefulness.

To elaborate the four dimensions, Liaw (2002) explained that "computer anxiety" is the fear of computers or the tendency of a person to be uneasy, apprehensive, and phobic towards current or future use of computers. "Computer

confidence" is the ability to use or learn about computers. "Computer liking" refers to enjoying working with computers. And "computer usefulness" is the degree of perceived usefulness of using computers for present and future work. However, computer anxiety and computer confidence have been suggested to be closely related or part of the same continuum (Kay, 1993; Nash and Moroz, 1997).

Loyd and Gressard's CAS has been and still is widely used by researchers and educators. For example, Mizrachi and Shoham (2004) used the first three dimensions (computer anxiety, confidence, and liking) to measure computer attitudes among Israeli B.Ed students. Similarly, Oosterwegel, Littleton, and Light (2004) also used the first three dimensions of the CAS to measure computer attitudes among children aged between 12-13, and van Braak (2004) measured self-perceived computer competence of university students with the computer confidence subscale of the CAS.

Kay (1993) developed the Computer Attitude Measure (CAM) in 1989 and revised it four years later. Kay defined computer attitudes as the persons' feelings of favorableness or unfavorableness towards computer and computer-related activities. The CAM measures four theoretically distinct constructs – cognitive, affective, behavioral, and perceived control. The cognitive component represents the perceptions of information about computers (e.g., stereotypical knowledge). The affective component represents one's feelings towards computers (e.g., fear, pleasure). The behavioral component represents behaviors consistent with the attitudes (e.g., avoidance of computer). Finally, the perceived control component refers to "the perceived ease or difficulty of performing a particular behavior" (Kay, 1993: 372).

Recently, Noyes and Garland (2005) reported the use of Kay's Computer Attitude Measure to measure students' attitudes towards books and computers in their study. Their questionnaire comprised the affective scales and response categories of the CAM developed by Kay in 1989. It included these 10 dimensions of affect: likeable/unlikable, good/bad, happy/unhappy, comfortable/uncomfortable, calm/tense, full/empty natural/artificial, exciting/dull, fresh/suffocating, and pleasant/unpleasant. Each dimension comprised a Likert scale with 7 points of agreement between the positive and negative dimensions. They argued that "it is debatable with some of the descriptors whether they are actually representing a positive and negative aspect" (p.

240). The participants in the study commented that the descriptions were difficult to apply. Noyes and Garland (2005) mentioned that the lack of reliability may arise from respondents having difficulty in applying the measure.

While the computer attitude scale of Todman and Dick (1993) comprises three subscales: fun, perceived usefulness and perceived ease of use, Woodrow (1994) used the four dimension scale of Loyd and Gressard (1984) plus three additional ones which are interest, gender, and acceptance.

Selwyn (1997) developed a tool to measure students' attitudes towards using and interacting with computers. The tools are based on four dimensions: affective component, perceived usefulness component, perceived control component, and behavioral component.

Levine and Donitsa-Schmidt (1998) used a computer attitude questionnaire which comprised 42 statements encompassing seven dimensions. They examined the relationships between the four of latent attitude dimensions (educational tool, tool of enjoyment, important tool, and stereotypes), a computer self-confidence, perceived computer knowledge, and three measures of computer use (at home, at school, and frequency of use).

More recently, Liaw (2002) measured attitudes towards computer and the Internet with two 16-item scales. Later, Yang and Lester (2003) conducted a Principal Components Analysis of responses to Liaw's attitude measure and found that although the computer attitude scale was internally consistent, the underlying constructs were less clear. The computer attitudes are more strongly associated with computer and Internet skills and behavioral components than with the Internet attitudes.

Though numerous and various definitions have been proposed, as seen in the literature review above, there is still no single, universally accepted definition of a construct for computer attitudes (Smith, Caputi, and Rawstorne, 2000). However, the theoretical framework suggested by Triandis (1971, cited in Noyes and Garland, 2005) proposed that an attitude made up of affective, cognitive, and behavioral components should be adopted in one of the more well-known measures of computer attitudes, the Computer Attitude Measure (CAM) and it is also adopted in the

construction of measures by many researchers (Kay, 1993; Noyes and Garland, 2005). Given this, Triandis's framework is adopted in this research.

4. The Interrelationships among the Three Variables

The interrelationships among the three individual variables have also been reported in many studies with mixed results. The relationships among the three variables are reviewed and presented below.

4.1 The Relationship between Computer Anxiety and Computer Familiarity

The relationship between computer anxiety and computer familiarity has been studied relatively widely. For example, Maurer (1994) suggested that computer-related experience seemed to have the most clearly direct relationship with computer anxiety. Yang, Mohamed, and Beyerbach in 1999 found that computer-related experience does influence computer anxiety. More specifically, the greater use of computer or the more computer experience was found to decrease anxiety (Levine and Donitsa-Schmidt, 1998, Yaghi and Abu-Saba, 1998). A meta-analysis by Chua et al. (1999) also confirmed an inverse association between computer experience and the level of computer anxiety. They mentioned that the correlation between computer anxiety and prior computer experience is the "most consistent finding."

Similarly, Bozionelos (2001a), based on previous research, stated that computer experience is the best measure for computer anxiety. Furthermore, Beckers and Schmidt (2003) mentioned that more recent findings seem to imply that actual experience with computers is a powerful determinant of computer anxiety.

Recently, Wilfong (2004) defined "computer use" as the number of hours an individual used a computer per week while "computer experience" as the mastery level in specific realms of computer software. He conducted a study to discover which factors are related to anger or anxiety of computer users. He concluded that both computer use and computer experience were all negatively correlated with computer anxiety.

No matter how many studies confirmed the relationship between the two variables, evidence has showed that computer experience does not always

decrease computer anxiety. Bozionelos (2001a) found that only the highly computer anxious individuals benefited most from treatment interventions based on provision of computer experience. The acquisition of computer experience was not very helpful for those who reported relatively low scores on computer anxiety. Furthermore, Gos (1996) revealed that the "quality" of prior experience was strongly related to a measure of computer anxiety and suggested that the computer experience should be a "positive and subjective" one.

4.2 The Relationship between Computer Anxiety and Computer Attitudes

The relationship between computer anxiety and computer attitudes has been reported in many studies. In some studies, computer anxiety was treated as a part of the constructs of computer attitudes. For example, Loyd and Gressard (1984) defined computer anxiety as one of the four parts of constructs of the computer attitudes. In constructing a questionnaire to survey the attitudes towards the computer, they included a subscale of computer anxiety into their Computer Attitude Scale (CAS). Another example is the study of Colley, Gale, and Harris (1994). They referred to computer attitudes as computer self-confidence, computer anxiety, and computer attitudes.

This perspective is contradictory to many studies and reports. For example, Heinssen et al. (1987) mentioned that computer anxiety should not be confused with negative attitudes towards the computer which entail beliefs and feelings about computers rather than one's emotional reaction towards using computers.

Computer anxiety and attitudes towards computers used to be seen as synonymous (i.e. an individual who experiences high levels of computer anxiety is said to have negative attitudes towards computers) or as separate concepts with the same antecedents, but the evidence suggests that computer anxiety is an intervening variable between other variables such as computer attitudes and demographics. Therefore, it appears that the two variables are distinct constructs (Shaft et al., 2004).

Though computer anxiety and negative attitudes towards computers are separated constructs, they are related (Whitley, 1997). While "computer anxiety"

involves a person's affective response when faced with computers, "computer attitudes" refers to a person's positive or negative beliefs about computers.

Based on the perspective that the two are distinct constructs, Bradley and Russell (1997) investigated the relationships of some components of computer anxiety among school teachers. One of their research results indicates a significant correlation between computer anxiety and computer attitudes.

More recently, Hong and Koh (2002) conducted a study with Malaysian rural secondary school teachers and found a negative linear correlation between computer anxiety and positive attitudes towards computers. They reported that teachers that had low computer anxiety also had positive attitudes towards computers.

4.3 The Relationship between Computer Attitudes and Computer Familiarity

Computer attitudes have been shown to correlate positively with computer experience. Busch (1995) found that one of the most important predictors of computer attitudes is previous computer experience. Similarly, Mizrachi and Shoham (2004) found computer experience produces positive computer attitudes. Levine and Donitsa-Schmidt (1998) found positive relationships between computer attitudes, computer use or objective experience, and confidence. Bozionelos (2001a) concluded that the more contact people have with computers, the more likely they are to express favorable attitudes regarding computers.

However, Shashaani (1994) found that the relationship between computer attitudes and experience is not straightforward, but rather it is dependent on the nature of the experience measure and the dimensions that are deemed component parts of the attitude construct. For example, it was found that hours of computer use each week and the number of courses taken are moderately associated with computer interest, while only a small magnitude of positive correlations is identified for either course enrollment and computer ownership with any of the attitude dimensions.

Although the relationships between computer attitudes, computer experience, and computer use have been found consistently, Garland and Noyes (2004: 836) mentioned quite clearly that many of the data reported in earlier literature violate

parametric assumptions. The result is that the magnitude of the relationship found can be overstated and is frequently a poor reflection of the "true" association.

In summary, the empirical evidence from previous studies shows that the interrelationships among the three computer-related variables are inconclusive. To achieve a better understanding of this issue, this study aims to investigate the interrelationships among the three variables to see if the relationships exist among them and, if they do exist, to investigate the magnitude of those relationships.

5. The Impact of the Computer Related Variables on Performances

A review of reports on the impact of the computer related variables on the performance of individuals is reviewed in this section.

Lee (1986) found that individuals with no computer experience (or only experience with game playing) received lower scores on a computerized test. A conflicting result was found by Mazzeo et al. (1991, cited in Eignor et al., 1998). In four out of five comparisons, computer experience was not related to performance on a computerized test. In one case in which a significant relationship was found, it was the opposite of what was expected in that higher computer familiarity was related to lower scores. Conversely, Taylor et al. (1999) concluded from their study, conducted with 1,204 examinees at 12 sites, that there was no evidence of adverse effects on the computer-based TOEFL performance due to lack of prior computer experience. They concluded that computer familiarity does not play a major role in the CBT TOEFL performance. Furthermore, Sawaki (2001: 44) in her literature reviewing study found that "the effect of examinees' characteristics, such as computer familiarity, does not seem to manifest itself in test scores".

Chou (2001: 9) studied the relationship between computer anxiety and learning performance and obtained different results with respect to the gender of the participants. The results from his study showed that female students with high computer anxiety improved their performance the best in the instruction condition, whereas high computer anxiety male students performed the worst in the same condition.

Fulcher (1999) found that attitudes towards taking computerized and paper-and-pencil tests have no significant effect on test scores. On the contrary, Russell

(1999) investigated students' performances on computerized and paper-and-pencil language tests and matched them with their self-reported preferred mode. He found that where self-reported preferences were matched to the test type, the performance was found to be higher.

In sum, the literature reviews have so far explored the areas of reading comprehension, language testing, and the test-takers' computer related variables. This information is employed in the construction of the tools in this study. The relationships among the three variables and the computer-based test scores presented in the literature reviews lead to the research questions of this study which is firstly whether those relationships are also present among Thai students and secondly whether those variables can be predictors for their reading comprehension CBT scores.

Research Approaches and Statistics

Hatch and Farhady (1982) listed different kinds of research such as historical, descriptive, and correlational research. Each kind of research needs different statistics to analyze the data. This study is a correlational research and aims to investigate the relationships between test-takers' variables and English reading comprehension ability. The statistics involved in this study includes basic, correlational, and predictive statistics. The basic statistics, which are mean, standard deviation, and range of scores, are used to explain the general description of the three variables. The studies using correlational and predictive statistics to investigate the relationships are presented in the following section.

1. Correlational Analysis

Isaac and Michael (1995: 46) mentioned that in correlational studies, researchers are interested in investigating the extent to which variations in one factor correspond with variations in one or more other factors based on correlation coefficients. Similarly, Hatch and Farhady (1982: 192) stated that in this kind of study, researchers determine the degree of the relationship between pairs of two or more variables.

The easiest way to visualize the relationship between the two sets of scores is to represent them graphically by plotting the values of one variable against the values of the other correspondingly (Hatch and Farhady, 1982; Hatch and Lazaraton, 1991). The plot is called a scatterplot or scattergram. If there is a positive perfect relationship of two variables, we can draw a line through the chart (Figure 2.1). It will be a perfectly straight line which refers to a perfect linear relationship.

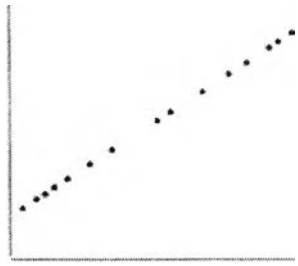

Figure 2.1 $r = +1$

Figure 2.2 $r = -1$

The problem is that to do just the scatter plotting does not give us any quantitative measure of the degree of linear relationship between the two variables. Thus, we also need to calculate the statistic r which is called *Pearson's product moment correlation coefficient* to describe the linear relationship between the two variables. It shows how closely the two variables are related. If there is a perfect positive relationship between the two sets of scores, the magnitude of the correlation coefficient would be +1 (Figure 2.1) and a perfect negative relationship would be -1 (Figure 2.2). Nevertheless, the perfect relationships do not happen in reality, not even when we measure the same thing twice (Hatch and Lazaraton, 1991).

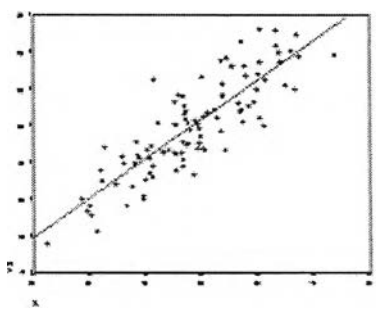

Figure 3.1 $r = .85$

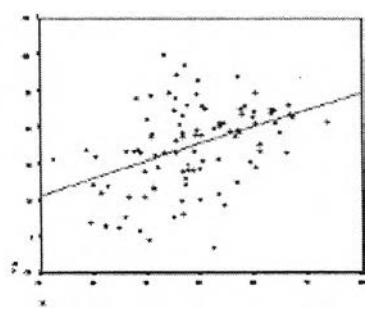

Figure 3.2 $r = .42$

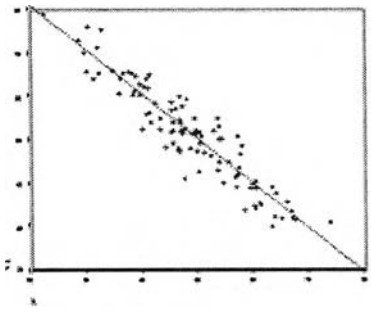

Figure 3.3 r = -.94

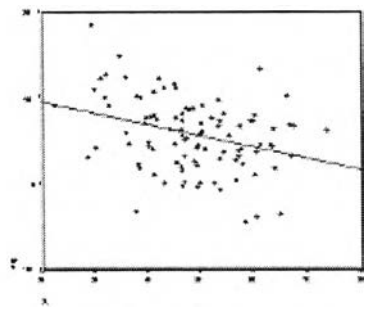

Figure 3.4. r = -.17

The greater the value r is, the stronger the relationship between the two variables. Thus, a correlation coefficient of r = .85 (Figure 3.1) indicates a stronger degree of linear relationship than r = .42 (Figure3.2). Likewise a correlation coefficient of r = - .94 (Figure 3.3) shows a greater degree of linear relationship than r = -.17 (Figure 3.4). A correlation coefficient of zero indicates the absence of a linear relationship between two variables (Figure 4.).

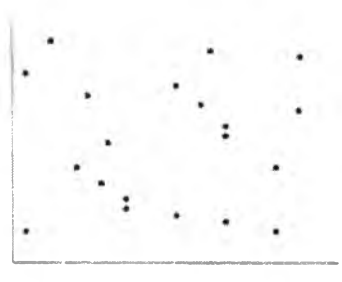

Figure 4. r = 0

In sum, the magnitude of the Pearson's product moment correlation coefficients can vary from +1 to −1. The greater the value, the stronger the relationship between the variables. A +1 correlation coefficient indicates the perfect positive correlation while a −1 correlation coefficient indicates the perfect negative correlation. On the other hand, a zero correlation coefficient indicates no relationship between the variables.

Mathematical procedures are available for testing the linearity of the relationship of the two variables (Tacq, 1997) but a scattergram is suggested because it is the easiest and most practical way to test and to visualize the linearity (Hatch and Farhady, 1982; Brown, 1990; Hatch and Lazaraton, 1991).

One of the basic assumptions that must be met before applying Pearson correlation as a measure of how well any two variables "go together" is that the relationship between *X* and *Y* must be linear. By linearity it means that it is possible to draw an imaginary straight line through the points on the scatterplot and measure how tightly the points cluster around that straight line (Brown, 1990).

However, linearity cannot be always assumed. Sometimes the line that connects the points is not straight, not linear but curvilinear. Brown (1990) gives graphical examples of curvilinear relationships in Figure 5. a-d and states that this kind of relationship is most frequently a problem when one of the variables is a function of time. Hopkins, Hopkins, and Glass (1996: 92) suggest that the variable "age" frequently has a curvilinear relationship with other variables. They also point out that poorly constructed tests can also give the appearance of curvilinear. It happens when the test is too easy (a "ceiling effect" where most students score highly) or if it is too difficult. However, this curvilinearity is spurious if the test does not demonstrate sufficient item difficulty or discrimination.

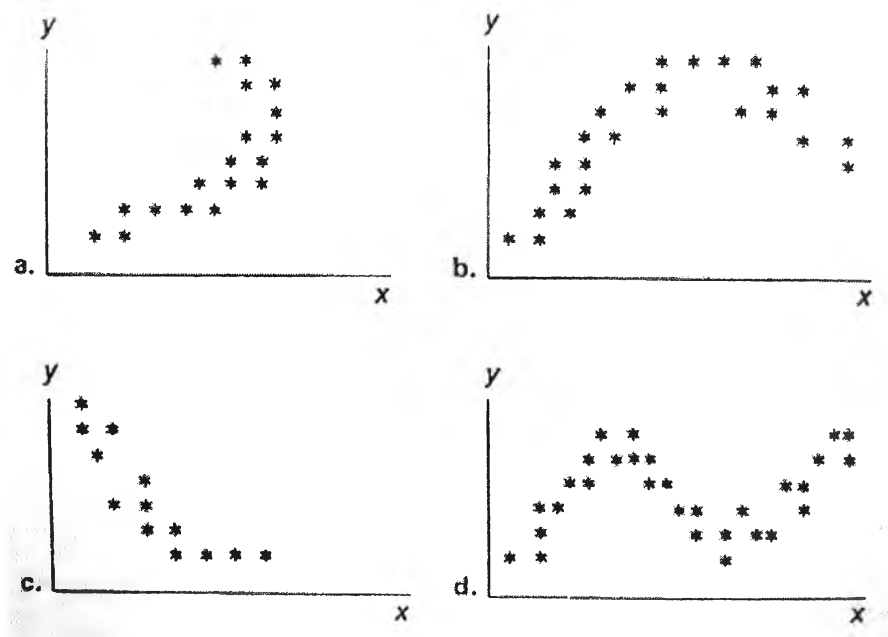

Figure 5. Curvilinears

In addition, Hatch and Lazaraton (1991: 437) give a classic example of the relationship between anxiety and test scores. They mentioned that we all need a little

anxiety because anxiety increases our performance. Nevertheless, they also pointed out that too much anxiety can have a very debilitating effect on the performance. As anxiety increases at the upper ends of the anxiety scale, scores on tests should decrease. The correlation starts out strongly positive and ends up strongly negative, and the actual *r* value would, of course, be distorted.

Furthermore, Hatch and Lazaraton (1991: 438) indicated that it is also possible to have a curvilinear relationship in the other direction. In teaching an introductory course in linguistics, many students have a difficult time comparing sound inventories and phonological rules across languages. In the early stages, the more students do their homework assignments, the more they are confused. The relation between studying time and success on phonology exercises may be negatively correlated. At a certain point, these same students magically "get" the concept, and from that point on, the time devoted to study is positively correlated to success on exercise scores.

Cohen, Manion, and Morrison (2000) gave some examples of bivariates that yield curvilinear relationships. They are

- Pressure from the teacher and students' achievement;
- Degree of challenge and students' achievement;
- Age and cognitive ability;
- Assertiveness and success;
- Pressure from principal and teacher performance;
- Age and concentration.

The Pearson correlation coefficient is, thus, not appropriate when the relationship is curvilinear because it violates the underlying assumption. Cohen, Manion, and Morrison (2000) suggested that in planning correlational research, researchers need to consider whether linearity is a reasonable assumption to make, or whether a curvilinear relationship is more appropriate.

However, Pedhazur (1997) mentioned that when the data depart from linearity, it is necessary to resort to non-linear models which can be classified into two categories which are firstly intrinsically linear models and secondly intrinsically nonlinear models. An intrinsically linear model is one that is linear in its parameters but nonlinear in the variables. Appropriate mathematical procedures need to be applied to transform nonlinear variables into a linear one. Intrinsically nonlinear

models, on the other hand, are nonlinear in the parameters and cannot be rendered linear by a transformation. Intrinsically nonlinear models are also called *essentially nonlinear* models (Fox, 1984). Accordingly, more sophisticated statistics for measuring the strength of curvilinear relationships are needed (Hatch and Lazaraton, 1991; Cohen, Manion, and Morrison, 2000).

In transforming the data, Edwards (1976) suggests a transformation of the X scale, the Y scale, or both the X and Y scales. He mentioned some simple nonlinear relationships that can be transformed into linear relationships which include the power curve, the exponential curve, and the logarithmic curve. Fox (1984) suggests a raising variable to power as an example of transformation. Suppose the equation $Y = a + bX$ is inadequate because the increases in X are attended by more-than-linear increases in Y as in Figure 6.a If all X-values are positive, the $X = X^2$ is a monotone transformation of X; then, the equation $Y = a + bX^2$ adequately represents the relationship between Y and X. Figure 6.b shows the "stretches" of the X-axis, affecting larger values of X more than smaller values.

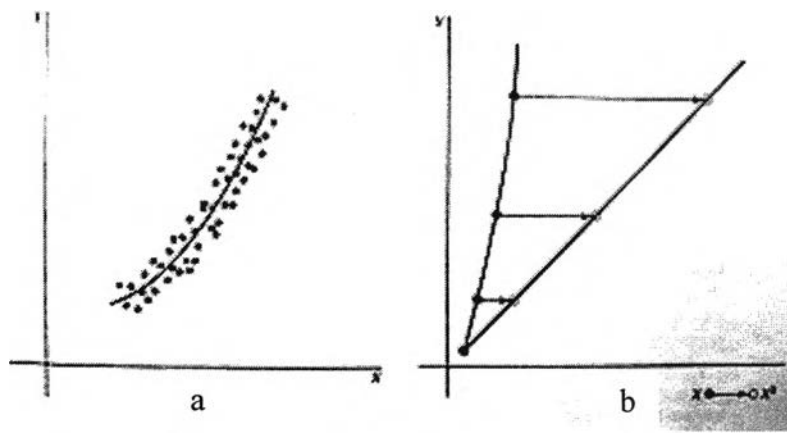

Figure 6. Straightening a Nonlinear Relationship

Besides raising the variable to powers, Pedhazur (1997) gives more examples of transformations which include expressing variables as logarithms and taking the square roots of variables.

In measuring the strength of intrinsically nonlinear relationships, Hatch and Lazaraton (1991) suggest that researchers should make an appointment with a statistical consultant in order to discuss appropriate statistical procedures.

Ezekiel and Fox (1959) suggest that the straight line is a type of relation of very great importance and usefulness. It is one of the simplest functions to fit and to explain and for that reason it is very widely used. Accordingly, this study which attempts to identify the relationships among test-takers' variables and the test scores of the students assumes the linear relationship and uses the Pearson product-moment correlation coefficient to identify the degree and the nature of the relationship.

When some of test-takers' variables do not linearly correlate with the test scores, the curvilinear should be identified whether the relationship is intrinsically linear or nonlinear. If it is linear in parameters but nonlinear in the variables, procedures such as making some adjustments on the tools (Hopkins, Hopkins, and Glass, 1996) or performing the transformations of data (Edwards, 1976; Hatch and Lazaraton, 1991; Cohen, Manion, and Morrison, 2000) should be employed. On the other hand, if it is nonlinear in the parameters and transformation cannot be rendered, the help from statisticians in measuring the strength of curvilinear relationships is required as suggested by Hatch and Lazaraton (1991).

If the test-takers' variables are not linearly correlated with the test scores, the nonlinear models are employed. Data transformations or sophisticated statistical procedures of the intrinsically nonlinear model are used to achieve a better understanding of the relationships among those variables. However, researchers need to be aware of certain issues in connection with this.

Wilcox (1996) elaborated on the advantages of the nonlinear model. He explained that the correlation between two variables plays an important role in statistics, but other pitfalls associated with the interpretation of r also warrant attention. He explained that if X and Y are independent random variables, it can be shown that $r = 0$. However, many take this to mean that when $r = 0$, X and Y are independent, but this is not necessarily the case. It is wrong to conclude that there is no relationship between X and Y when $r = 0$. Figure 7 shows some points that lie on a half circle given by the equation $Y = \sqrt{25^2 - (X - 25)^2}$. He said that not only are X and Y dependent but there is also an exact relationship between X and Y in the sense that, if you are given a value for X, then Y is determined exactly.

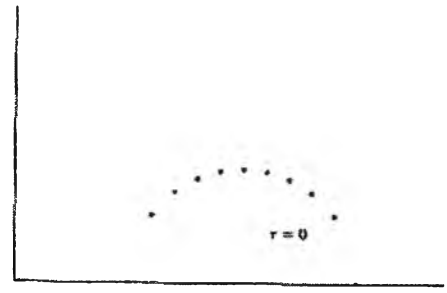

Figure 7. A Half Circle Relationship

Thus, the use of data transformations, which stretches and straightens the curve, provides more accurate conclusions and interpretations about the relationship of the two variables.

Nevertheless, Garson (2004) mentioned that pre-processing step nonlinear transformation of selected variables may run the danger of *overfitting* the model to what are, in fact, chance variations in the data. He suggested that power and other forms of transformations should be done only if there is a theoretical reason to do so because it runs the risk of introducing multicollinearity in the regression model.

Tacq (1997) explained that multicollinearity occurs when the independents are highly intercorrelated, $r > .6$. The first independent variable that is entered into the equation will assume a disproportionate amount of the explanatory power which leads the researcher to conclude wrongly that the subsequent variables are causally unimportant or type II error. However, Garson (2004) pointed out that the problem of multicollinarity can be handled in many ways, e.g. increasing the sample size, the use of centering, removing the most intercorrelated variables from the analysis, etc.

In conclusion, researchers should employ the nonlinear procedures; e.g. the transformation of data with cautions because they can provide better understanding and more accurate interpretation of the relationships. However, at the same time, they can in certain circumstances also mislead researchers who should be aware of such a risk.

Examples of correlational research studies include Khine's (2001) and Nash and Moroz's (1997). Khine (2001) investigated the intercorrelations of the computer attitude subscales while Nash and Moroz (1997) used the Pearson formula to find the correlations between the subscales of the Computer Attitude Scale and computer activities at home and at work.

2. Multiple Regression

Hatch and Farhady (1982: 233) mentioned that multiple regression is an important procedure to identify which independent variable is "more important or contributes more to the dependent variable". It gives the combination of variables which predict performance on the dependent variable (Hatch and Lazaraton, 1991). Some researchers use the term "predictor variables" for "independent variables" and the term "criterion variables" for "dependent variables".

Predictor variables can be entered into the multiple regression equation by the "hierarchical" method and the "statistical" method. In the hierarchical method or the "enter" method, researchers determine the order of entry of the variables based on theories of previous studies. In the statistical method, the order in which the predictor variables are entered is determined by the computer. This statistical method has several versions which are "forward" selection, "backward" selection, and "stepwise" selection (Hatch and Farhady, 1982; Brace, Kemp, and Snelgar, 2000).

In the forward selection, the independent variables are entered into the model one at a time according to the strength of their correlations with the dependent variable which is the one that is significant and with the highest beta. While each variable is entered the effect of adding is assessed. If the variable does not significantly add to the success of the model, then it is deleted until no more independent variables that are significant (Hatch and Farhady, 1982; Brace et al., 2000).

In the backward selection, all of the independent variables are entered in the equation. The variables with no significant parameters are removed and the regression is re-calculated. If the removal of the weakest variable significantly weakens the model then the independent variable is re-entered; otherwise it is removed. The process is repeated until only useful predictor variables remain in the model (Brace et al., 2000; Miles and Shevlin, 2001).

Stepwise is the most sophisticated statistical method. This technique is a combination of the backward and forward techniques which adds variables when they are significant, and removes them when they are not significant (Miles and Shevlin, 2001). The process begins when an independent variable which is best correlated with the dependent variable is entered in the equation and its value is assessed. If it

contributes to the model then it is retained. Then the remaining independent variable with the highest partial correlation with the dependent variable is entered, controlling the previously-entered independents variable. The process is repeated until adding one more independent variable does not significantly increase the R-square. Therefore, only useful predictor variables remain in the model and the smallest possible set of predictor variables is provided (Brace et al., 2000; Garson, 2004).

Brace et al. (2000) mentioned that if you have no theoretical model in mind, and/or relatively low numbers of cases, then it is probably safest to use the "enter" method. He mentioned that statistical methods should be used with caution and only when a large number of cases are included because "minor variations in the data due to sampling errors can have a large effect on the order in which variables are entered and therefore the likelihood of them being retained" (Brace et al., 2000: 211). Researchers should therefore use each technique of the multiple regression with extreme caution (Miles and Shevlin, 2001).

Chan (2004: 60) mentioned that the multicolinearity analysis and the residual analysis should be carried out before accepting the final model of the multiple regression analysis.

"Multicollinearity" exists when the independent variables are strongly intercorrelated. Only the first variable included in the equation will have a significantly higher contribution than the other variables that follow. Interpreting the results from variables entered in the equation in such an order can be mistaken (Hatch and Farhady, 1982; Hatch and Lazaraton, 1991). Garson (2004: 7) mentioned that high R^2 increases the standard error of the beta coefficients and makes the assessment of the unique effect of each independent variable either difficult or impossible. To assess multivariate multicollinearity, he suggested the use Tolerance or VIF. He suggested that the problem with multicollinearity is indicated when either Tolerance is less than .20 or VIF is more than 4.

Tacq (1997) mentioned that the purpose of residual analysis is to additionally test for outliers, linearity, and homoscedasticity. Outliers are cases with very high residuals. They are the clear exceptions to the regression explanation. Linearity is an assumption of regression analysis. When nonlinear relationships are present then conventional regression analysis will underestimate the relationship. In addition,

homoscedasticity is the condition when the residuals are not dispersed randomly throughout the range of the estimated dependent. In other words, the variance of residual error should be constant for all values of the independents.

Garson (2004: 10) mentioned that outliers exist when the standardized residual is greater than 3.3 (corresponding to the .001 alpha level). He also pointed out that **partial regression plot** is used to assess nonlinearity. It simply plots each independent on the x axis against the dependent on the y axis. The preferred method for assessing nonlinearity is **partial residual plot** which is a visual form of the t-test and the b coefficient. It shows a given independent on the y axis and the corresponding partial residual on the x axis. Finally, the **simple residual plot** shows both nonlinearity and non-homoscedasticity (heteroscedasticity). When points form a cloud with no trend or a random pattern, then there is no nonlinearity or heteroscedasticity. Nonlinearity is shown when points form a curve. Non-normality is shown when points are not equally above and below the y axis 0 line. Lastly, non-homoscedasticity is shown when points form a funnel or other shapes showing the variance which differs as one moves along the y axis.

Up until this point, the definition, entry methods, and statistics procedures of multiple regression analysis have been explored briefly. The following paragraph explores some studies that employed this multivariate statistics in investigating the independent variables which are similar to ones of this study.

In the studies of the relationships among test-talkers' variables, many researchers reported their uses of multiple regression. Shermis and Lombard (1998) examined the degree to which computer and test anxiety had a predictive role in performance across three computer-administered placement tests. Bradley and Russell (1997) investigated the intercorrelations of respondents' characteristics and computing competence, anxiety, and specific sources of anxiety as well as the predictors of computer competence and anxiety using multiple regression analysis. Karsten and Roth (1998) studied the relationship of computer experience and computer self-efficacy to performance in an introductory computer literacy course by using both correlational and multiple regression analyses.

In this study, multiple regression provides two things:

1. the most appropriate equation to predict reading comprehension CBT scores of the students, and

2. how many and which independent variables should be brought to calculate the predictive reading comprehension CBT scores of the students.

In conclusion, the literature review on research approaches and statistics provides the details of the analysis procedures and interpretation of the results for this study. The results yielded from the analysis of data are intended to answer the research questions about the relationships among the three variables and the reading comprehension CBT scores.

Summary

Chapter Two reviews studies that mainly on four areas that include the reading theory, the computer-based testing, the computer related variables, and finally the research approach and statistics. This literature review provides the basis on the instrument construction, data collection and analysis, and interpretation of results for this study. Furthermore, the relationships among the three computer related test-taker variables and the computer-based test scores reported in the literature review also lead to the questions of this study.

The computer related test-takers' variables were studied at the university level in several countries and they have made important contributions to the area of language testing. However, there is no evidence of any investigation on the computer related variables of Thai students at university level in Thailand. Accordingly, study explores these issues with the synthesis of the theoretical and empirical perspectives from previous studies and applies the correlational research design in an attempt to find the answers to the research questions poised in Chapter One.