CHAPTER II

LITERATURE REVIEW

This chapter reviews the past literature concerning five main areas related to the research study. The first part involves the scholars' explanation of the listening ability and the listening comprehension processes. The second part provides an overview of learning strategies that have been discussed by researchers. The third covers the area of listening comprehension assessment. The fourth reviews the issues concerning the integration of tasks in language tests. Lastly, the views and theories about computer-based assessment are provided.

2.1 Listening ability and listening comprehension processes

2.1.1 What is listening ability?

According to Kannika Kreutanu (1998), listening ability means the ability to comprehend the meaning of words, phrases, sentences and messages as well as to identify general ideas and main ideas. Moreover, it also includes the competence of drawing conclusions as well as using prior and new knowledge to decode the meaning correctly as conveyed by the speaker.

Definitions provided by other researchers as quoted by Kannika Kreutanu (1998) are as follows:

According to Nunan (1991), listening ability is the listener's ability to combine all of the information in one unit and interpret it by using his or her background knowledge.

Sarunya Sriwichai (1994) referred to the listening ability as the ability to use one's English knowledge and background knowledge, together with his or her new knowledge of grammatical structures to draw the meaning of the listening input conveyed by the speaker.

Other definitions introduced by academics emphasized that listening comprehension is really a process, a very complex process. For example, Peterson (2001) referred to listening comprehension as a multilevel, interactive process of meaning creation. Buck (2001: 114) termed second language listening ability as "the ability to process extended samples of realistic spoken language, automatically and in real time, to understand the linguistic information that is unequivocally included in the text, and to make whatever inferences are unambiguously implicated by the content of the passage." Similarly, Duzer (1997) perceived that listening comprehension is an 'active' process of choosing and interpreting an input from both auditory and visual cues despite the fact that it is categorized as a passive skill.

2.1.2 The processing of listening comprehension (bottom-up and top-down)

Two main types of processing that explain how one understands a listening input are the 'bottom-up' and 'top-down' approaches. The 'bottom-up' approach suggests that to understand a verbal input, the smallest sound segments that are meaningful, or phonemes, are initially decoded. After that, words are identified, followed by sentences or the decoding at the syntactic level. This processing continues to the next level which is the analysis of the semantic content, and then the literal meaning is interpreted based on the communicative situation. On the other hand, the 'top-down' approach emphasizes the opposite steps of decoding an acoustic input (Buck, 2001).

Brindley (1997) discussed the same processing of listening comprehension. According to Brindley, in the bottom-up processing, the smallest unit is identified. Then, listeners will put those smaller units together to form a larger unit, and deduce the meaning from those structures. Top-down processing, on the other hand, depends on the use of context to comprehend the input.

Hadley (2000) also explained similar second language comprehension processing which involves learners' knowledge of the target language code (linguistic knowledge) and knowledge of the world (cognitive familiarity). In addition, she put emphasis on knowledge of discourse structure or students' ability to understand

different types of discourse, e.g. conversations, literary texts, political speeches, etc. In relation to cognitive processes, Hadley discussed how bottom-up and top-down processing come into play. The first was referred to as data-driven processing since the process moves from part to whole while the latter was viewed as conceptually driven since the process begins with a wider scope and background knowledge is used in prediction.

The bottom-up and top-down processes can also be explained through two forms of knowledge: linguistic and non-linguistic. Linguistic knowledge refers to the knowledge concerning phonology, lexicon, semantics and syntax whereas non-linguistic knowledge means students' existing knowledge of the world or their background knowledge. It is believed that top-down processing is employed when learners cannot depend on their linguistic knowledge. In such a manner, the whole context will be focused to facilitate comprehension, and non-linguistic knowledge plays a significant role. On the contrary, the role of non-linguistic knowledge or cognitive familiarity is less crucial when students use the bottom-up processing in which linguistic knowledge is more heavily in focus (Yi'an, 1998).

Which mode is activated during the stage of listening comprehension? Brindley (1997) quoted two different views towards these two types of processing. Kelly (1991) claimed that the bottom-up processing is the strategy used by learners at an early stage of learning. The top-down manner is employed by learners at a later stage. This claim was argued by Richards (1988) who proposed an opposite view. He made a strong argument that learners who have low proficiency do not have much linguistic knowledge to rely on. This is why non-linguistic knowledge or top-down processing is used.

Buck's (2001) argument supports Richards' view. He claimed that listening comprehension truly involves the 'top-down' process since various genres of knowledge must be used. The knowledge is not employed in a linear manner; on the contrary, it sometimes occurs simultaneously in an interactive way.

In addition to the top-down and bottom-up approaches, Buck (2001) claimed that mental models can represent how texts are processed. Van Dijk and Kintsch (1983) in Buck explained that while discourse is being processed, the listener constructs two things, namely, a text base and a situation model. The first is a representation of semantics (concept and what describes it). Keeping the text base or propositions is truly a burden. In processing texts, we usually remember the overall events described in the discourse rather than isolate linguistic items. The latter way to process texts, which is the most common, is called a mental model. We remember only the meaning of what we have heard. This is also called a situation model. It explains how we store implicit ideas that are not overtly conveyed, and how the information is added and adjusted as well as waiting to be matched with a new piece of information.

Duzer (1997) mentioned the two processes just like other researchers and underscored the importance of bringing learners to be aware of the effects of the application of these two processes on their listening comprehension. She suggested that they must be given opportunities to use both types of processing to enhance their listening ability.

In conclusion, it is agreed among researchers that the bottom-up and the top-down modes are used by learners to draw meaning out of the 'received' input. The researchers also tend to agree on the domination of the roles of the two approaches by claiming that either will be more relied on depending on the tasks involved as well as the proficiency level of the listeners.

2.1.3 What do learners do when they listen?

Duzer (1997: 2), in her work, proposed clear and systematic basic processes of listening at work. They are as follows:

1. determining a reason for listening
2. taking the raw speech and depositing an image of it in short-term memory

3. trying to organize the information by identifying the type of speech event (conversation, lecture, radio ad) and the function of the message (persuade, inform, request)
4. predicting information expected to be included in the message
5. recalling background information (schemata) to help interpret the message
6. assigning a meaning to the message
7. checking that the message has been understood
8. determining the information to be held in long-term memory
9. deleting the original form of the message that had been received into short-term memory

According to the researcher, these processes do not have to occur one after another, but can happen simultaneously in rapid succession. They can occur either backwards or forwards depending on the listeners, but the listeners usually are not aware of performing the steps.

When taking the researchers' views towards the listening process into consideration, it is very obvious that all kinds of processing involve listeners' cognitive stages. Therefore, studies in their use of cognitive strategies have been pursued widely in hopes that they will unveil the mysteries of how the human brain processes an oral input that will lead to comprehension and how teachers and educators should teach and implement these strategies in their classroom.

2.1.4 Factors affecting listening comprehension

The factors which are believed to influence one's listening comprehension, whether enhancing or degrading one's listening ability, are numerous. Bachman (1990) proposed a model presenting sources of variation in language test scores. There are four main factors, namely, communicative language ability, test method facets, personal characteristics, and random factors.

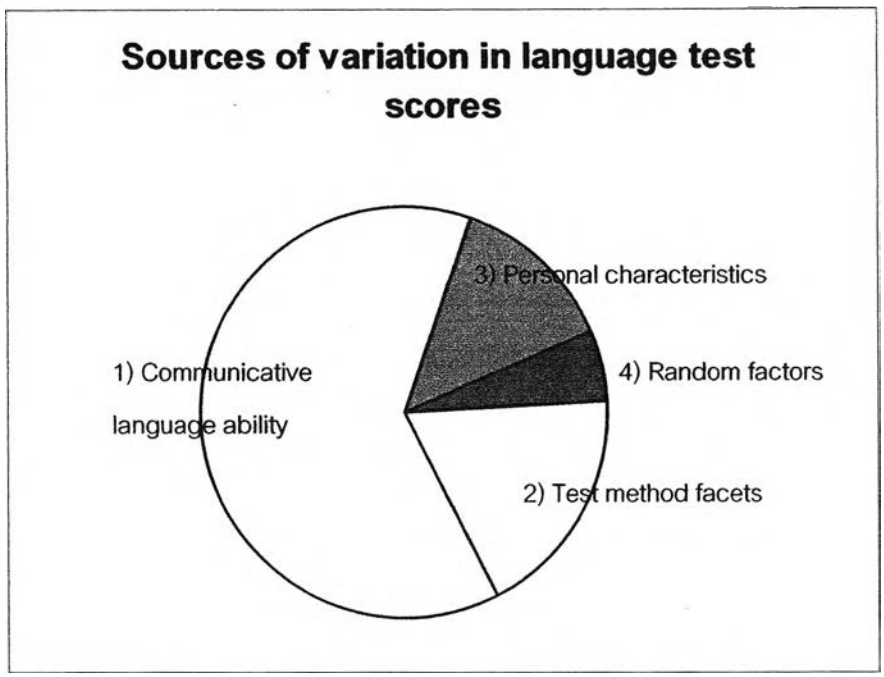

Figure 2.1: Sources of variation in language test scores

(Bachman, 1990: 350)

The first factor that is communicative language ability consists of language competencies, strategic competence and psychophysiological mechanisms. Language competencies mean the test takers' abilities to appropriately execute their competence when communicating in various situations. They refer to test takers' knowledge in various areas including their grammatical competence, textual competence, illocutionary competence, and sociolinguistic competence. The grammatical competence is the test takers' knowledge of vocabulary, morphology, syntax, and phonology. The textual competence involves the knowledge of understanding cohesion and rhetorical organization. It is related to test takers' understanding of semantic connections of language elements, and to their knowledge of how the information is organized, e.g. narration, description, classification, etc. Regarding the illocutionary competence, it is related to the theory of speech acts. Test takers must be capable of understanding any underlying meanings of the conveyed message. For example, a sentence like "It's cold in here", may directly express the literal meaning or it may also convey someone's request to turn on the heater (Bachman, 1990). The knowledge in this area enables test takers to use language to meet various functions. Lastly, the sociolinguistic competence

refers to test takers' ability to understand a variety of language uses such as different dialects and registers.

Strategic competence is the second factor underlying communicative language ability. It consists of assessment, planning and execution components. Firstly, test takers must possess the ability to evaluate the input, the functions of the input, and the context involving the input. Secondly, they are able to plan which kind of knowledge, such as grammatical knowledge, is needed to lead them to achieve successful communication.

The last component in communicative language ability is psychophysiological mechanisms. It involves test takers' capability to distinguish one mode and channel from another, and also to effectively make use of various modes and channels simultaneously. For example, they are able to rely on both visual and auditory channels at the same time, or they can deal with both receptive and productive modes simultaneously.

The second factor, which is test method facets, includes various components: environment, rubric, facets of input and expected response, and the relationships between input and response. First of all, testing environment refers to locations, equipment and physical conditions. Secondly, test rubric consists of test organization, time allocation and instructions. It involves factors that determine how test takers are expected to take a test. Thirdly, the facets of the input and the expected response are concerned with the format and the nature of language that are used in the input, or expected to be used in the output. The format is, for example, the speed of the input, the channel of presentation or the language accounts for test takers' performance. The last component is relationships between input and response. There are three types of input and response that are reciprocal, nonreciprocal and adaptive. Reciprocal input and response in tests requires test takers to interact with the input, and probably feedback is given such as in oral interviews. Nonreciprocal tests involve no interaction and feedback such as cloze tests and dictations. Adaptive tests refer to those require interaction but

lack feedback. For example, adaptive tests are tailored to test takers' responses to previous question items.

The third factor that is believed to have an effect on test takers' scores is personal characteristics that include cultural background, background knowledge, cognitive abilities, sex and age. These components vary according to each individual.

Lastly, there are random factors that are generated by interactions of the three factors (communicative language ability, test method facets and personal characteristics) and measurement error. The measurement error refers to variance that has no direct relationship with the objectives of tests such as environment, administration procedures and test items.

The communicative language ability addressed above agrees with what Brindley (1997) discusses as lexical knowledge and syntactic knowledge. Lexical knowledge, especially the key lexis, is required to prevent miscommunication or communication breakdown. Syntactic knowledge also accounts for listening ability in that it can increase the amount of content retained in short term memory. That means if students possess certain syntactic knowledge of the target language, they will be able to understand and keep what they have heard in their short term memory.

Memory is another factor that is obviously important in language comprehension. Memory is negatively affected depending on the difficulty of tasks. Yi'an (1998) proposed that comprehension is related to limitations of human memory capacity. A task that requires a high cognitive demand such as a listening comprehension test can negatively impact the memory capacity. The process will slow down and some elements will be forgotten. This can account for why students understand everything in the verbal input but forget it at the end, resulting in poor performance on the test. Other factors are speech rate and noise.

The attitude towards listening tasks as being very demanding is also mentioned in Hadley (2000). Yorio (1971), quoted by Hadley said that listeners will forget the cues they receive from the input very easily because they must simultaneously predict the future cues and make associations with the past cues. The process can be painful to novice learners of a second or foreign language. In other words, listeners have to concentrate on a three-part process, which includes storage of the past cues, prediction of future cues, and associations between the two.

Apart from memory capacity, speakers' rates of delivery and accents also count. Accent is a factor enhancing or reducing listeners' comprehension. When listeners, even native speakers, hear a new accent for the first time, they must tune their ears in order to understand the message. It is also the same with non-native speakers who are more familiar with standard English. For non-native speakers, such accents as the working-class accent of New Yorkers or foreign-accented English can have a negative effect on language learners' comprehension. Similarly, for native speakers of English, Gass and Selinker (1994) revealed that native speakers' comprehension of non-native speakers' utterances increased when they could cling to non-native speakers' vocabulary and pronunciation. Despite non-native speakers' good grammatical sentences, the native speakers judge them as poor if they have problems with the pronunciation.

These two examples interestingly illustrate how phonological features, namely accents and pronunciation (e.g. stress) impact the comprehension of native and non-native speakers of English. Although accents deviant from standard have been considered a difficult factor by language learners, in their study, Ross and Langille (1997) insisted the incorporation of various non-native accents of English in a listening test since it would enhance the authenticity of the listening assessment. The importance of making teachers and students aware of the non-native accents will be discussed further in the 'English as an International Language' (EIL) section.

Like Brindley (1997), Duzer (1997) put background knowledge as an important factor for comprehending spoken messages. A familiar content can ease the degree of

difficulty of the message since listeners have been exposed to the vocabulary and possess background knowledge of the topic of the conversation before.

This view is also supported by Carrell (1981, 1983, 1984,1987), who claimed that background knowledge was an important factor in ESL/EFL comprehension. This is because it is directly related to the process of decoding meaning as mentioned earlier that background knowledge, which is regarded as non-linguistic knowledge, is what listeners rely on in the top-down approach (Yi'an, 1998).

The notion of 'background knowledge' is defined by various researchers such as Lingzhu (2003) who defined background knowledge as prior knowledge. It's the knowledge that students use to relate to the content of the input they read or listen to. Holmes (2002) stated that background knowledge is the existing knowledge of the world. Nieh (2004) defines background knowledge or prior knowledge as the knowledge that a reader or a listener already possesses.

In fact, background knowledge is a term that covers a wide range of knowledge. Carrell and Eisterhold (1988) quoted in Kannika Kreutanu (1998) divided background knowledge into two genres.

1. Background knowledge that concerns the structures of messages; for example, in narrating stories (expository), there are certain structures governing the narration e.g. setting, beginning, development, climax and ending.

2. Background knowledge that deals with content schematic knowledge such as fields of study. It is believed that those who are familiar with those topics will be able to comprehend input easier.

Background knowledge is significant not only when communicating in a discourse, but also when one is trying to decode what he or she hears. In the first situation, when conversing, if two interlocutors share different background knowledge or

world view, miscommunication can take place. On the other hand, if they share the same knowledge, they can understand each other by means of inferences (Buck, 2001).

Why is background knowledge so important? Apart from serving as non-linguistic knowledge that students activate when using a top-down approach, it is part of the schema theory that explains the process of comprehension. Buck (2001) defined schemata as structures representing knowledge in memory. It sets the interpretation of text and expectations for people, places or events. For receptive skills, input is overlaid by the pre-existing knowledge in an attempt to find a match. Therefore, students' existing knowledge or background knowledge plays a part in comprehension (Lingzhu, 2003).

Like other factors that can either enhance more comprehension or make the input more difficult, background knowledge can be viewed as a source of bias. When writing a test, teachers must consider to what extent the content of the text is familiar and relevant to the experience of certain groups of students (Holmes, 2002). Bachman (1990) agreed that the interaction between students' knowledge of content area and their performance in listening and reading tests do exist. He pointed out that it is difficult to distinguish students' language proficiency from their background knowledge, especially in English as a Specific Purpose (ESP). Therefore, to make sure that the interpretation of scores is correct, it must be identified at the outset what we are going to measure: the content or the language ability.

The last two factors that can accommodate or downgrade learners' comprehension of real-life conversations, if the input is deprived of them, are redundancy and paraphrase. They are characteristics of spoken language. They help modify discourse since both interlocutors can negotiate. However, rarely do these features of interactionally modified discourse appear in second language listening tests adding to the degree of difficulty of the test as well as harming the construct validity of most language listening assessment.

2.1.5 English as an International Language

Talebinezhad and Aliakbari (2001) simply defined English as an International Language (EIL) as the use of English by people across the world for communication. EIL is an intervarietal way of communication. From EIL perspective no speaker is realized as extreme. English is not viewed as a property of its native speaker, but it has become the world's property.

According to Jenkins (2003), the roles of English can be characterized as a language used as one's native language, second language, and foreign language. This implies a variety of English used by people around the world. This is based on the grounds that there is no one variety of English from one territory to another and from one region within one territory to another. Moreover, a large group of native speakers have been living in countries where English is spoken as a second language such as in India, Hong Kong and where it is used as a foreign language, such as in Thailand.

The use of English as a lingua franca in various kinds of businesses, international trade, diplomacy and tourism is widely found. This can explain why people are learning English more than any other language (Smith, 1983, cited in Talebinezhad and Aliakbari, 2001). According to the information taken from the English-speaking Union of the Commonwealth (1998-2002), it is hard to gain official figures of the population speaking English. However, it can be estimated that 377 million people speak English as their first language whereas around one-third of the world's population speak English as a second language. The number of people who speak English as a foreign language accounts for a wide range from 300 million to 750 million. The figures can only be treated as approximate since it is impossible to quantify the large number of people learning English at school.

Due to the fact, there has been a new orientation towards teaching EIL, many authors such as Jenkins (2000) and MacKay (2002) published books concerning various aspects of this genre of English, namely the Phonology of EIL and Teaching EIL, respectively.

Why should EIL be recognized and learned? Talebinezhad and Aliakbari (2001) cited Smith (1983) that studying EIL is important and speakers whose English is a mother tongue should study EIL if they expect to communicate with non-native speakers or native speakers who use a different national variety. And due to various uses of English in each culture, people should be helped to use EIL in communication. Perhaps, teaching or exposing students to EIL such as accents is crucial since students will gain an opportunity to hear actual spoken English with its inherited diversity in everyday life (Baxter, 1980, cited in Talebinezhad and Aliakbari, 2001).

2.1.5.1 EIL in language assessment

With regard to EIL in language assessment and evaluation, Jenkins (2003) revealed that so far there are few signs of including non-native speaker variation and innovation into language assessment despite the fact that the number of speakers who are not native speakers of English exceeds the number of those whose mother tongue is English. In Jenkins (2003), Lowenberg (2000) said that World English testing, then, still reflects very strongly the 'deficit linguistics' view strictly attached to native speakers' norms. Kachru (1992) proposed a solution to the inappropriateness of English language testing around the world. He said there should be a clear distinction defined whether English is used in monolingual or multilingual societies as he called a 'paradigm shift'.

As much more awareness has been raised concerning English as an International Language, teachers as well as test developers should start thinking about the way to assess authentic language use. Probably, Kachru's (1992) suggestion of defining what is to be tested is one of the best ways to cope with the current situation.

2.1.5.2 EIL accents in listening tests

If a significant quality of a good test is authenticity, then should we integrate, for example, EIL features into the test? As for a listening test, one of the important aspects is accent. Ross and Langille (1997) claimed that general listening tests are not designed with the assumption that such factors as different accents are involved. They also stated

that a test that claims to assess international English must possess certain linguistic features that are used by both native and non-native speakers such as the interlanguage phonology.

To make an important move concerning this issue, factors concerning EIL accents that can trigger problems must be considered. For example, Buck (2001) claimed that even for native speakers, they cannot completely understand new accents, but they will do more when they get used to it. This is also the same with non-native speakers, who usually do not comprehend non-standard accents, especially in colloquial speeches. Major, Fitzmaurice, Bunta et al (2002) reported the effects of nonnative accents on listening comprehension of both nonnative and native speakers of English. They scored significantly lower when taking listening comprehension tests that contained nonnative accents of English. However, with some limitations, they insisted that further investigation be conducted since this may depend on the degree of foreign accent of the speakers who delivered the talks on the tests.

As the notion concerning EIL has been well recognized in language teaching, language assessment must also move in accordance with the concept. If an authentic language test is the ultimate goal, accented English must, then, be included. The problem is how to create such a test with that particular quality and without any threats to validity and reliability. Perhaps, more research studies should be conducted to explore if there are any effects of the interlanguage accent on student performance, and how to eliminate them.

2.2 An overview of learning strategies

This part deals with the framework of cognitive and metacognitive strategies proposed by various researchers. In addition, the studies that explore the benefits of the two strategies are discussed.

2.2.1 Definitions of learning strategies

Recent researchers have been interested in learning strategies since it is believed that these strategies affect how learners manage their own processes of language acquisition, as well as their language use in social interactions and on tests. Therefore, learning strategies definitely account for one's success or failure in learning a new language.

In Carson and Longhini (2002), Oxford's (1990) definition of learning strategies was quoted as specific actions employed by learners in an attempt to make their learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to new situations. Based on Oxford (1990), there are two types of learning strategies, namely direct and indirect strategies.

The direct strategies are related to the target language of learners. They are listed as follows:

1. Memory strategies – these strategies are used by learners to store and retrieve information required to perform a language task.
2. Cognitive strategies – these strategies are used when learners want to select what to pay attention to so that understanding will be enhanced.
3. Compensation strategies – these are the strategies that learners use to make up their lack of knowledge in certain areas, which obstruct their understanding.
4. Conversation strategies – these strategies were proposed and added by Carson and Longhini (2002). They are used when learners initiate repairs and requests for assistance.

On the other hand, the indirect strategies are those which support a second language acquisition. They include:

1. Metacognitive strategies – these refer to the strategies that regulate the cognitive processes.
2. Affective strategies – these strategies are related to self-regulated attitudinal and emotional factors, which affect one's new language learning.

3. Social strategies – these strategies are chosen by learners to increase their understanding and improve their production of the language being learned by means of interaction.

The types of strategies listed above correspond to those reflected in the test, "Strategy Inventory for Language Learning (SILL)" developed by Oxford (1990). In the test, strategies are divided under the following categories: remembering effectively (memory), using mental processes (cognitive), compensating for missing knowledge (compensation), organizing and evaluating (metacognitive), managing emotions (affective), and learning with others (social).

Research studies have shown none of the strategies out of those listed above suits all language learning contexts. Griffiths and Parr (2001) reported different strategies are used in relation to diverse theories or language methodology. For example, memory and cognitive strategies are involved in the development of vocabulary and grammatical knowledge (grammar translation) while compensation and social strategies support the communicative competence theory and teaching approach. Also, in their study, the strategies taken from the SILL were used to explore the gap between teachers' perception of learners' use of strategies and that of learners' towards their own usage. The learners reported that they used metacognitive strategies more than cognitive strategies, whereas the teachers perceived that the learners relied on cognitive strategies more than metacognitive strategies.

2.2.2 Definitions of cognitive and metacognitive strategies

Among the strategies proposed by researchers, cognitive and metacognitive strategies were mostly discussed. Najar's (1998) descriptions of successful language learners can be explicitly linked to their use of cognitive and metacognitive strategies.

Purpura (1999) quoted many definitions in his work. Generally, researchers defined cognitive strategies as 'actions or behaviors' that learners invoke during language learning, language use or language testing. For example, O'Malley and

Chamot (1990) explained that cognitive strategies are the strategies that learners rely upon while 'manipulating or transforming' an input so that comprehension will occur. Anderson (1982, 1985) said cognitive strategies are related to information processing. They are a set of 'behaviors' and involve 'mental manipulations or transformations' of materials of tasks. Wenden (1991) also referred to cognitive strategies as those relating to 'mental' steps or operations employed by learners when processing both linguistic and sociolinguistic inputs.

In his own research, Purpura (1999: 7) adopted Wenden's (1991) definition claiming that cognitive strategies, in his view, are "a set of strategies or processes which are related to behaviors associated with each stage of the learning processes; however, instead of limiting cognitive strategy use to learning. Cognitive strategies are also viable behaviors in language use and testing contexts. Therefore, cognitive strategy use is a set of conscious or unconscious mental or behavioral activities or operations which are directly or indirectly related to the comprehending, storing or retrieval of information during language acquisition, use or testing"

Metacognitive strategies are considered self-management. Theoretically, they have certain influence on cognitive strategies. The examples of these strategies are planning for, monitoring and evaluating a learning event. In Purpura (1999), according to O'Malley and Chamot (1990), metacognitive processing refers to learners' thought and their knowledge of the learning process. They plan for their learning, monitor their learning while performing a task, and evaluate themselves after accomplishing the task. Wenden (1991) also used the term 'self-management' to define metacognitive strategies. The meaning of the strategies is extended by some researchers such as Faerch and Kasper (1983) and Bachman and Palmer (1996). The first two researchers included the strategies learners use when communication breaks down into the scope of metacognitive strategies. Other than the planning and assessment stages, Bachman and Palmer (1996) viewed goal setting as a part of the strategies as well.

2.2.3 Cognitive and metacognitive strategies and language learning

According to Najar (1998), many research studies revealed that successful learners are usually those who are active and are responsible for their learning. They know how to deal with their learning situations such as studying a textbook when preparing for a test. They are also aware of how to regulate their learning processes such as setting a specific goal that they will successfully pass a test. In some work such as in Song (2005) reported the positive effects of the use of certain strategies such as linking with prior knowledge strategy and monitoring strategy on test takers' language performance on the Michigan English Language Assessment Battery (MELAB) test.

Interestingly, and supported by Oxford (1990), not only do successful learners possess various kinds of learning strategies, but they must also be able to select appropriate strategies for each situation. For example, those who highlight main ideas and take notes can remember the content more than those who do not when performing a reading task. Highlighting and taking notes are parts of a number of strategies used by learners. They also vary according to the skills being activated. For instance, strategies, used in listening comprehension can be different from those employed in reading comprehension. Najar's (1998) examples of learning strategies, such as reading a textbook in advance to prepare for an examination, highlighting the main keys, keeping up with their notes and setting their objectives or success, can be categorized into cognitive and metacognitive strategies.

Learners' greater use of strategies is evidenced in the work of other researchers. For example, Aek Phakiti (2003) found that successful achievers used metacognitive strategies significantly more than unsuccessful test takers. Moreover, he found the existence of a significant relationship of test takers' use of cognitive and metacognitive strategies and their performance on a reading test. Like the results found in Aek Phakiti, those in Liu (2004) revealed that the more proficient students make use of learning strategies more frequently than the less proficient ones.

Successful learners are also reported as those who make an appropriate use of strategies. Their knowledge of strategies is not as important as how effective those learners use them (Anderson, 1985; Abraham and Vann, 1987). They are able to take advantages of a wider range of strategies compared to unsuccessful learners (Chamot, Küpper and Impink-Hernandez, 1988 in Purpura, 1999).

However, Vinther's work (2005), in which the use of cognitive strategies in learning syntax via a computer-based program was studied showed fewer cognitive strategies used once learners approached nearly automatic level of performing the tasks. The study stressed that the number of cognitive strategies that were verbalized through the verbal protocol decreased when students could master the language tasks.

2.2.4 Cognitive strategies and listening performance

Cognitive strategies, which are related to one's mental processes and regarded as direct strategies, have been claimed as significant for one's listening performance. Purpura (1999) emphasized that most recent models and theories involving SLA acknowledged the importance of cognitive processes. This view is supported by many researchers e.g. Derry and Murphy (1979), Oxford (1990), Chittiya Suwaphab (1998), Rubin, Quinn and Enos (1998), etc.

Benefits of cognitive strategies used in listening comprehension can be found in the past literature. For example, Chittiya Suwaphab (1998) advocated that the learner strategies believed to help enhance students' listening abilities were cognitive and metaconitive strategies. In the same way, other researchers such as Oxford (1990) revealed that learners who were trained to use cognitive strategies would be able to build their listening skill. According to Oxford (1990) the cognitive strategies that involve learners' listening performance are as follows:

1. Practice – it is regarded as the most important among all of the cognitive strategies. It consists of five more strategies.

1.1 Repeating or imitating will make learners familiar with the verbal messages and their components, e.g. vocabulary, expressions, etc. and will promote their understanding.

1.2 Practicing with sounds and writing systems will enhance learners' understanding of tone, intonation, etc.

1.3 Recognizing formulas and patterns will help learners capture and remember phrases or sentences frequently used, e.g. 'how are you?', 'I'm fine thank you and you?'

1.4 Recombining meaningful phrases or small chunks is the strategy that learners use to combine the small chunks of language that they know to understand the whole sentences. For example, learners understand the meaning of 'weather's fine.', 'like to' and 'walk' they should be able to understand the meaning of the whole sentence.

1.5 Practicing naturalistically will bring all learners to be exposed to authentic tasks, e.g. following directions and advice, role play, etc.

2. Receiving and sending messages

2.1 Getting the idea quickly includes such strategies as previewing question, pictures, charts, symbols or others will be used in previewing.

2.2 Using resources for receiving and sending messages can be done through the use of dictionaries, or if on a test, the use of context will be relied on.

3. Analyzing and reasoning

3.1 Reasoning deductively means learners apply the rules that they have learned before their listening activities. For example, hearing a question beginning with an auxiliary, they will be able to guess that the answer should be introduced by 'yes' or 'no'.

3.2 Analyzing expressions refers to how learners try to understand the language by isolating words, phrases or sentences. The best example is from the words that contain prefixes and suffixes. If well taught, students will be able to use the strategies for better comprehension.

3.3 Analyzing contrastively is used when learners want to compare every component of the target language to their mother tongue.

3.4 Translation is believed to benefit the beginning stage of learning since the sentences and vocabulary are not very complex.

3.5 Transferring relies on learners' use of their prior knowledge. This prior knowledge can refer to that obtained through their language learning processes or their experience.

4. Creating structure for input and output

4.1 Taking notes will allow learners to keep important information. And it is recommended that the notes should be taken in the target language and be well organized.

4.2 Summarizing requires more thinking than taking notes. The input must be precise and capture all the conveyed main ideas.

4.3 Highlighting is not a complicated strategy. Learners can just highlight important information, use capitalization or bold print or circle it.

Apart from Oxford (1990), Derry and Murphy (1979) as well as Thompson and Rubin (1996), cited in Chittiya Suwapap (1998) cited five cognitive strategies underlying the listening performance as follows. Their listening cognitive strategies are comparable to those of Oxford (1990).

1. predicting content
2. listening to the known (either familiar or only partial familiar words)
3. listening for redundancies (one of the characteristics of spoken languages)
4. listening for tone, voice and intonation (also one of the spoken language characteristics)
5. resourcing (drawing meaning from contexts)

Duzer (1997) is among numerous researchers who explored learners' listening processes. She proposed three processes which can be regarded as cognitive strategies and one process, which can be categorized as a metacognitive strategy.

Cognitive Strategies:

1. When listening to input, learners must determine a reason for listening, take the raw speech and deposit an image of it in short-term memory.

2. They must attempt to organize the information by identifying the type of speech event (conference or lecture) and determine the function of the message (persuasion, announcement, or request).

3. Learners' schemata or background knowledge is drawn on to help interpret the message.

Metacognitive Strategies:

Learners check if the message has been understood and determine if the information should be kept in their long-term memory.

To conclude the definitions and the descriptions of cognitive strategies proposed by the researchers, the cognitive strategies used in performing a listening task are summarized in Table 1 below. It shows the listening cognitive strategies based on the listening strategy frameworks of Oxford (1990) in comparison with Purpura (1999).

Table 2.1

A Comparison of Cognitive Strategies Used in Performing a Listening Task by Purpura (1999) and Oxford (1990)

Purpura (1999)	Oxford (1990)
1. Comprehending processes	
- Analyzing contrastively	- Analyzing contrastively
- Analyzing inductively	-----
- Clarifying or Verifying	- Using resources
- Inferencing	- Getting the ideas quickly (previewing)
- Translating	- Translating

Table 2.1 (Continued)

A Comparison of Cognitive Strategies Used in Performing a Listening Task

by Purpura (1999) and Oxford (1990)

2. Storing or memory processes	
- Associating	- Recombining meaningful phrases
- Linking with prior knowledge	- Transferring
- Repeating or rehearsing	- Repeating or imitating
	- Practicing with sounds/ writing systems
- Summarizing	- Summarizing
3. Retrieval processes	
- Applying rules	- Reasoning deductively
	- Recognizing formulas and patterns
	- Analyzing expressions
- Practicing naturalistically	- Practicing naturalistically
- Transferring	- Transferring

According to Table 2.1, the classification of the strategies and the terminology describing the strategies that the two researchers used are both similar and different. They can be grouped as shown in the table so that the picture of the comparison can be clearly presented under the three main processes: comprehending processes, storing or memory processes and retrieval processes.

The first strategy which is related to the comprehending processes consists of five sub-processes that are analyzing contrastively, analyzing inductively, clarifying or verifying, inferencing and translating. The analyzing contrastively strategy is the strategy that is used when learners attempt to compare any elements of the target language such

as phonology, lexicon, syntax to those of their first language. It also refers to their comparison of the elements in the same language such as minimal pairs. The analyzing inductively strategy, according to Purpura (1999) is the process that describes learners' attempts to understand the input by analyzing it and trying to make generalizations from the input. Oxford's (1990) framework does not propose any strategy comparable to this one. The clarifying or verifying strategy means learners try to understand by asking themselves, asking others, or consulting any sources for further clarification or verification. It is similar to Oxford's using resources strategy. Like Purpura's inferencing strategy, Oxford's previewing strategy refers to learners' attempt to understand the input by guessing words, predicting outcomes, interpreting from the message that is not directly conveyed, and determining the writer's or the speaker's attitudes or tones. Lastly, translating from the target language to their first language is agreed by both researchers as one of the strategies learners use to enhance their comprehension.

The second strategy which is related to storing or memory processes consists of four sub-strategies that are associating, linking with prior knowledge, repeating or rehearsing and summarizing. The associating strategy is used when learners try to remember an input by categorizing it into meaningful categories and make a semantic link between or among elements. Oxford (1990) also presented similar strategy that is recombining meaningful phrases in her work. The second sub-strategy, linking with prior knowledge, is claimed by Purpura (1999) as learners' referring to their prior knowledge or the information that they know such as knowledge concerning grammatical rules, vocabulary, topics. Based on Oxford's proposal, a similar strategy that can be compared to Purpura's linking with prior knowledge strategy is transferring. Repeating is another strategy that both researchers regard as helpful for input retention. Lastly, summarizing is also mentioned by both researchers as an approach to memorize input. According to Purpura, it can be done through any form: oral, written or mental.

The third cognitive strategy involves the retrieval processes that include three sub-strategies: applying rules, practicing naturalistically and transferring. The description of applying rules strategy is straightforward. It means learners try to refer to

the rules that they have learned. This is the same as Oxford's (1990) reasoning deductively strategy and recognizing formulas and patterns. Although analyzing expressions may not be explicitly linked to the applying rules strategy, its process which involves the use of knowledge of affixes can be considered as rule application. Practicing naturalistically is another strategy claimed by both researchers as effective as it can lead to better performance. This strategy includes practices such as following directions or advice, role playing, and communicating with others. The last sub-strategy is transferring that seems to overlap the use of prior knowledge. This refers to learners' attempt to retrieve the knowledge they have to enhance their comprehension.

Two strategies taken from Oxford's (1990) framework that are not listed in Purpura's (1999) are note-taking and highlighting. Perhaps this is because Purpura's framework is based on cognitive 'processes' referring to learners' mental processes, while note-taking and highlighting are physical strategies that are used to enhance the cognitive processes. This is supported by the definition of cognitive strategies given by Wenden (1991) who referred to cognitive strategies as those relating to 'mental' steps or operations employed by learners when processing both linguistic and sociolinguistic input. In other words, they are adopted by learners to make important information more salient.

And although no strategies provided by Oxford (1990) can 'directly' be matched with Purpura's (1999) analyzing inductively' strategy, several strategies can be 'associated' to it such as analyzing contrastively, using resources, translating, etc, since these strategies involve the processes of analyzing input, formulating hypotheses and making generalizations (analyzing inductively).

2.2.5 Metacognitive strategies and listening performance

Metacognitive knowledge consists of knowledge that focuses on or regulates any single cognitive activity (Flavel, 1985, cited in Kasper, 1997). According to O'Malley and Chamot (1990), metacognitive processing is learners' thoughts and their knowledge of the learning processes. This includes planning for and monitoring their learning (while

performing a task) as well as evaluating themselves (after accomplishing the tasl Wenden (1991) referred to metacognitive processes as self-management. Bachman and Palmer (1996) viewed goal setting as a part of the strategies as well.

An example of research studies concerning the use of metacognitive strategies is from Millman, Bishop and Ebel (1965), who proposed several tips for taking tests. Most of them are found related to metacognitive strategies, including setting up a mental schedule for progressing through the test (planning).

Brown (2002) introduced various useful strategies for language learners. It is obvious that he gave importance to metacognitive strategies since he stated clearly that learners should plan their own goals. Based on his questionnaires concerning strategies for taking tests, he brought learners to focus on all stages of test taking: before a test, during the test, and after the test. These steps are also reflected in the listening strategies for a test proposed by NCLRC (2004: 2). The metacognitive strategies are:

1. Before listening: Plan for the listening task
   1.1 Set a purpose or decide in advance what to listen for
   1.2 Decide if more linguistic or background knowledge is needed
   1.3 Determine whether to enter the text from the top down or from the bottom up

2. During and after listening: Monitor comprehension
   2.1 Verify predictions and check for inaccurate guesses
   2.2 Decide what is and is not important to understand
   2.3 Listen/view again to check comprehension
   2.4 Ask for help

3. After listening: Evaluate comprehension and strategy use
   3.1 Evaluate comprehension in a particular task or area
   3.2 Evaluate overall progress in listening and in particular types of listening tasks
   3.3 Decide if the strategies used are appropriate for the purpose and for the task
   3.4 Modify strategies if necessary

The two frameworks are compared to show the similar view they share towards the metacognitive strategies.

Table 2.2

A Comparison of Metacognitive Strategies Used in Performing a Listening Task by Purpura (1999) and NCLRC (2004)

Purpura (1999)	NCLRC (2004)
1. Goal-setting processes	
- Setting goals	- Setting a purpose or deciding in advance what to listen to
2. Assessment processes	
- Assessing the situation	- Deciding if more linguistic or background knowledge is needed - Determining whether to enter the top-down or the bottom-up approach
- Monitoring	- Verifying predictions and checking for accurate guesses - Deciding what is and is not important to understand - Listening/viewing again to check comprehension
- Evaluating	- Evaluating comprehension in a particular task or area - Evaluating overall progress in listening and in particular types of listening tasks - Deciding if the strategies used were appropriate for the purpose and for the task - Modifying strategies if necessary

Table 2.2 (Continued)

A Comparison of Metacognitive Strategies Used in Performing a Listening Task by Purpura (1999) and NCLRC (2004)

Purpura (1999)	NCLRC (2004)
3. Planning processes	
- Formulating a plan	- Setting a purpose or deciding in advance what to listen for - Deciding if more linguistic or background knowledge is needed - Determining whether to enter the top-down or the bottom-up approach
- Learning to learn	-----

The first strategy that is related to goal setting processes is agreed by Purpura (1999) and NCLRC (2004) as one of the metacognitive strategies that learners use to increase their understanding. Learners set goals of what to pay attention to; for example, they plan to listen to the words that they have to fill in the blanks.

The second metacognitive strategy is associated with assessment processes. There are three primary sub-strategies, which are assessing the situation, monitoring, and evaluating. According to Purpura (1999), assessing the situation refers to how learners assess the knowledge they possess, the context that they are dealing with and also other constraints. NCLRC (2004) also reported the same type of strategy although different terms were given. They described it as learners' decision if more linguistic knowledge is required or their determination of which approach is to be used. Monitoring is the second sub-strategy proposed by Purpura. This includes learners' monitoring their own performance on a task. Those comparable strategies proposed by NCLRC are described as learners' verifying their predictions and their guesses, their decision of

what is not important for comprehension, and their listening or viewing the input again to check their comprehension. The third sub-strategy, evaluating, which is the term suggested by Purpura, can also be matched with the same term proposed by NCLRC.

The planning processes consist of two sub-strategies: formulating a plan and learning to learn. Formulating a plan, which is the strategy proposed by Purpura (1999) refers to learners' generation of an overall plan of action before doing a task or their change of plan while doing a task. This is similar to what NCLRC (2004) proposed as setting a purpose or deciding in advance what to listen to, deciding if more linguistic or background knowledge is needed, and determining the approaches to be used. These three strategies proposed by NCLRC also overlap other strategies Purpura suggested. Since the third process, 'formulating a plan', of Purpura (1999) is similar to all of the 'before listening strategies' proposed by NCLRC (2004), the latter are comparable to 'formulating a plan'. Regarding 'learning to learn', which refers to learners' attempts to arrange the conditions to assist them to achieve a language task, it can be related to the whole process of metacognitive strategies. Researchers agree that these metacognitive strategies have certain effects on learners' performance.

2.2.6 Learning strategies and computer-based tests

Test-taking strategies are the test-taking processes which the respondents have chosen and which they are conscious of, at least to some degree (VAC, 2004). Do the strategies used in taking a paper and pencil listening test differ from those used in a computer-based listening test? A work in progress entitled 'What Do Listeners Do in a Computer-based Listening Test?' presented at the 23rd Annual Language Testing Research Colloquium (Kim, Kim and Shin, 2001) emphasized that no research had been done on the effects of multimedia and test takers' listening strategies on the computer-based test performance. In their study, they intended to investigate test takers' listening strategies and examine whether these strategies contributed to construct-relevant or-irrelevant variance. Also, they would look at the relationship between those strategies and learners' test performance. Although the target language focused in their study is Korean, the result is expected to be valuable to the assessment field.

In general, research studies reported how to incorporate learner cognitive and metacognitive strategies into computer-assisted language learning (Soo, 1999; Hsiao, 2004). Available research concerning learning strategies and computer-based programs proposed the same learner cognitive and metacognitive strategies discussed earlier since almost all of the questions asked still required the same learning strategies, e.g. inferencing, analyzing deductively, summarizing, etc. However, if the physical elements of a test were changed; for example, a computer-based test may incorporate more visuals via pictures, charts or video, there would be a potential that these changes would impact more usage of one strategy over another such as previewing. This is why test developers must be careful when they write a computer-based test. They have to ensure the validity and reliability of the program interface design.

One research study reported the difference of the strategies used between two types of TOEFL tests: paper-pencil and computer-based. It revealed that the biggest difference in these two tests was the strategy for guessing. However, 'Guessing' here refers to random selection of the choices, rather than the cognitive strategy of using context for guessing meaning. Random guessing to finish the listening section can hurt their scores since it is a computer adaptive test. Therefore, rushing to finish up all questions within the given time period will move them to easier questions and finally result in lower scores (Phillips, 2003).

Another study aimed to develop a CALL program to promote greater awareness of the strategies available and to help students discover strategies which would benefit them. In this study, the strategies incorporated included metacognitive strategies (strategy planning, self-monitoring and self evaluation), cognitive strategies (resourcing, note taking, summarization, grouping, deduction, inferencing, substitution, translation and transfer) and social strategies (questioning and cooperation). The cognitive and metacognitive strategies studied reflected the same strategies used in paper and pencil tests, which were presented earlier in the paper (Bull, 1997).

In summary, the cognitive and metacognitive strategies expected to affect learners' language performance lay within those frameworks presented by Purpura (1999) and other researchers. According to the literature in the past, the strategies used in a computer-based listening test have not been widely explored and no significant differences have yet been predicted.

2.2.7 Research concerning cognitive and metacognitive strategies

As we have learned, cognitive and metacognitive strategies play important roles in being successful language learners, or even test takers. There have been many research studies conducted on the two learner strategies. A work carried out by Sheorey and Mokhtari (2001), who studied the differences in the metacognitive awareness of reading strategies among native and non-native speakers of English, revealed that both groups were aware of almost all of the strategies used, according to their self-report. Moreover, from the survey, it was found that they attributed the same order of importance to categories and reading strategies. In both groups, the high ability-students reported higher usage of both cognitive and metacognitive strategies to help them interpret the reading input.

Another work done on reading strategies belongs to Purpura (1999). He found the relationships between cognitive strategies and metacognitive strategies. In other words, he discovered that metacognitive strategies did have positive impacts on cognitive strategies although they did not have direct effects on test-taker performance. It was interesting to learn that cognitive processes had no significant effects on reading and grammar abilities of the test-takers; however, the memory or storing processes did have a negative effect on them. This was because the whole process involved drawing information from long-term memory. Spending too much time to learn or remember can negatively affect the rate of recall.

Kasper (1997) studied a link between metacognitive knowledge and writing performance. The purpose of her study was to develop ways to enhance student writing competence, which was a major challenge to them. The results revealed that it was

necessary to design activities that guided students to use writing strategies. And these activities, along with instruction that supported students' use of metacognitive strategies, should be introduced at the beginning and as an integral part of the writing instruction.

Regarding the listening skill, Hale and Courtney (1994) found the effects of note-taking and student listening comprehension performance on the TOEFL test. They claimed that taking notes had little effect on student performance while encouraging students to take notes significantly impaired their performance. Despite the results showing the uselessness of note-taking, this study's scope was limited to only monologues or mini-talks. More studies on the longer talks like academic lectures or academic discussions needed to be pursued to see if the results are comparable.

Although the studies of learner cognitive and metacognitive strategies have been widely in focus for a long time, Purpura (1999) still suggested more future research on language skills other than the reading skill that he had already explored. Moreover, he also advocated a change of some variables such as the test used in the study. Studies on the differences between the use of strategies between different nationalities, gender and age were also suggested.

2.3 Assessing listening abilities

This part reviews the areas concerning listening comprehension assessment such as the approaches to create a listening test, the framework of the listening constructs, and other elements to be considered in English listening assessment.

2.3.1 Why is it important to study the listening processing?

Littlewood (1981) stated that listening is the skill we use the most. It is less controllable than the speaking skill since we cannot select the language that we will use. What we can do is to try to extract meanings as much as we can. To understand the message, we have to match our receptive repertoire with our own productive repertoire as well as cope with a wide range of situational and performance variables that are

beyond our control. Oranoot Chirdchoo and Jirada Wudthayagorn (2001) reported Thai students' opinions towards the listening skill in the same way. The students felt that the listening skill was the most difficult of all because they were not able to control the message. Littlewood (1981) added the reasons why the skill is considered difficult.

Firstly, the situations where speeches are delivered can distract the message such as background noise, distance or unclear sound reproduction (e.g. over loudspeakers at airports or stations). Secondly, the speeches may not be well planned, containing false starts and hesitations, which are characteristics of everyday spoken language. Thirdly, speakers may vary in terms of tempo of speech, clarity of articulation and regional accent, especially English, since it is often used by both native and non-native speakers. Considering the reasons behind the views and the facts towards the listening skill, we as language teachers must carefully take these threats into consideration.

Not only do listeners or students have problems with the skill, teachers and test developers do also. Listening involves cognitive processes which are difficult to assess and describe. To begin with, the process of understanding input should be explored before moving towards listening assessment.

Therefore, it is crucial for test developers to realize that listening comprehension is far more complex than a process of decoding. Meaning is not simply extracted, but is constructed through the active inferencing and hypothesis building process.

2.3.2 Approaches to listening assessment

Buck (2001) explained clearly in his book the three approaches of assessing listening skill that have been used by teachers.

2.3.2.1 Discrete-point approach

By using this approach, test developers evaluate language proficiency in isolated elements. This is influenced by the audio-lingual method that was once popular

among language teachers. As for listening assessment, Lado (1961), quoted in Buck (2001), recommended each constituent of language to be evaluated, namely segmental phonemes, stress, intonation, grammatical structure and vocabulary. The types of questions suggested were multiple-choice questions, true/false and pictures. Too much use of context was cautioned against. The importance of context was only for resolving any ambiguity.

However, despite many followers, this approach was attacked due to different opinions towards language competency. Oller (1983) defined the discrete-point test as a test that evaluates a single particular segment of language at a time. The problem raised is the threat to validity, which leads to a negative impact on the reliability; the reliability is meaningless without the validity. As a result, an alternative to the discrete-point tests, an integrative test, has been used so that the goal of assessing language competency that includes the use of language in context can be reached. Well-known examples of integrative testing are, for example, cloze and dictation.

2.3.2.2 Integrative testing

Oller (1979), cited in Buck (2001) tried to test the ability to use elements of knowledge at the same time. He emphasized the importance of context and pragmatic by defining that particular test as a task that makes learners process series of language components conforming to the normal contextual constraints governed by the rule of that language. He stressed the use of language, which contradicts Lado's (1961) emphasis of one's knowledge of a language.

2.3.2.3 Communicative testing

This kind of approach followed the communicative teaching trend. Communicative competence or the evaluation of one's use of language to communicate in real-life tasks has been in focus. Canale and Swain (1980) defined communicative competence as including the knowledge of grammar, sociolinguistics and strategy in using language to communicate. Bachman and Palmer (1987) said that it is one's grammatical and sociolinguistic competence.

The problems with these types of measurement are the generalization of students' ability on test tasks into those tasks in the real world, and the difficulty in creating those tasks. The reason behind the problem concerning generalization is that there is a wide range of communicative situations and contexts. It can be argued that the tasks that represent samples of those situations do not reflect the whole domain of language. The idea was advocated by Bachman (2003) and suggested by Baker (1989) that sampling the language elements may work better than extracting the target language use (TLU) domains since situational contexts where language is used are too various to be effectively sampled.

Therefore, choosing one method of assessment over the others is not a simple task. One approach will probably be more advantageous than the others. Taking the purpose of the assessment into consideration can lead to sound decision and proper solutions to problems that may arise.

2.3.3 Listening constructs

To assess students' listening ability, which is a measurable product of their cognitive processing, teachers or administrators must make sure that the test can elicit those particular constructs and that they can lead to logical interpretation of students' ability.

In general English language listening assessments, test takers' listening ability can be interpreted from the operations they perform during a test such as distinguishing sounds, listening for main ideas, making inferences, understanding concepts, etc. (Weir, 1993: 98-99). A list of operations (listening comprehension) that is included when testing general listening ability as suggested by Weir (1993) is as follows:

1. Direct meaning comprehension:

- listening for gist: This construct measures test takers' ability to understand general ideas of an auditory input.

- listening for main idea(s) or important information: This construct measures test takers' ability to trace the development of an argument, distinguish the main idea(s) from supporting details, differentiate a statement from examples, differentiate a proposition from its argument, and distinguish fact from opinion when clearly marked.
- listening for specifics: This construct measures test takers' ability to identify important details required for answering questions.
- determining speaker's attitudes and intentions: This construct measures test takers' ability to determine the attitudes and the intention of the speaker where it is obvious from the input.

2. Inferred meaning comprehension:
- making inferences and deductions: This construct measures test takers' ability to infer or make generalization of the input.
- relating utterances to the social and situational context: This construct measures test takers' ability to understand the social and situational contexts of the input.
- recognizing the communicative function of utterances: This construct measures test takers' ability to understand the function or the purpose of the input.
- deducing meaning of unfamiliar lexical items from context: This construct measures test takers' ability to guess the meaning of unfamiliar vocabulary by using the context.

3. Contributory meaning comprehension (microlinguistic):
- understanding phonological features (stress, intonation, etc.): This construct measures test takers' ability to understand all phonological elements such as stress, intonation so that they can distinguish similar features.
- understanding concepts (grammatical notions) such as comparison, cause, result, degree, purpose: This construct measures test takers' ability to understand the input by using their grammatical knowledge. For example, they are able to understand that 'er' is a marker for a comparative and it conveys certain meanings.
- understanding discourse markers: This construct measures test takers' ability to understand discourse markers such as transitions and conjunctions, especially the meaning that they express when combining two or more discourses.

- understanding syntactic structure of the sentence and clause such as elements of clause structure, noun and verb modification, negation: This construct measures test takers' ability to understand how sentences or clauses are formed. It assesses their knowledge of syntax or structures of sentences, clauses and phrases.
- understanding grammatical cohesion: This construct measures test takers' ability to understand the relationships between parts of the input through grammatical cohesion devices such as pronoun references.
- understanding lexical cohesion through lexical set membership and collocation: This construct measures test takers' ability to understand lexical chains or lexical semantic relations that are formed or perceived between words.
- understanding lexis: This construct measures test takers' knowledge of vocabulary.

4. Listening and writing (note taking from lectures, telephone, conversations, etc.):
- ability to extract salient points to summarize the whole text, reducing what is heard to an outline of the main points and important details: This construct measures test takers' ability to draw important points from what they hear and summarize the input.
- ability to extract selectively relevant key points from a text on a specific idea or topic: This construct measures test takers' ability to choose key words or phrases from the input.

Some classify these operations into two main groups: higher and lower order skills. The performance that requires understanding of clearly stated information, e.g. listening for the gist or main idea is lower-level processing, while those which require drawing inferences are regarded as higher order skills (Brindley, 1997). He adapted the taxonomy of listening skills from Weir (1993) and Rost (1990) and proposed his framework as presented below.

1. Orienting oneself to a spoken text:
- identifying the purpose or genre of a spoken text: This construct measures test takers' ability to identify the purpose or type of the input; for example, whether it is a persuasion, argumentation or fact presentation.

- identifying the topic: This construct measures test takers' ability to state the topic of the input.

- identifying the broad roles and relationships of the participants (e.g. superior/ subordinate): This construct measures test takers' ability to recognize the roles of interlocutors taking part in conversations.

2. Identifying the main idea(s) in a spoken text:

- distinguishing main ideas from supporting details: This construct measures test takers' ability to separate the main ideas of the spoken text from its specific details or other supporting details.

- distinguishing facts from examples: This construct measures test takers' ability to differentiate a statement from examples.

- distinguishing facts from opinions when explicitly stated in text: This construct measures test takers' ability to differentiate facts from opinions.

3. Extracting specific information from a spoken text:

- extracting key details explicitly stated in text: This construct measures test takers' ability to elicit main points that the spoken text conveys.

- identifying key vocabulary items: This construct measures test takers' ability to recognize the key vocabulary in the input.

4. Understanding discourse structure and organization:

- following discourse structure: This construct measures test takers' ability to understand the structures of discourses and successfully follow the discourses.

- identifying key discourse/cohesive markers: This construct measures test takers' ability to identify how discourses are related to one another, whether as dependent, complementary or reinforcing.

- tracing the development of an argument: This construct measures test takers' ability to follow an argument presented in an input.

5. Understanding meaning not explicitly stated:

- relating utterances to the social/situational context: This construct measures test takers' ability to understand the social and situational contexts of the input.
- identifying speakers' attitudes/emotional state: This construct measures test takers' ability to determine the attitudes and the emotion of the speaker where it is obvious from the input.
- recognizing the communicative function of stress/intonation patterns: This construct measures test takers' ability to understand the function or the purpose of the input by recognizing the stress and intonation that are used by the speaker.
- recognizing the speaker's illocutionary intent: This construct measures test takers' ability to recognize the real purpose of an auditory input. It is related to 'speech acts' whose meanings can be different according to situations or contexts.
- deducing meaning of unfamiliar words: This construct measures test takers' ability to guess the meaning of unfamiliar vocabulary by using the context.
- evaluating the adequacy of the information provided: This construct measures test takers' ability to assess whether the information presented is sufficient or not.
- using information from the discourse to make a reasonable prediction: This construct measures test takers' ability to make use of the information provided for logical prediction.

Others such as Buck (2001) proposed simple communicative listening constructs as presented below. Also, Buck (2001: 105) presented listening constructs which were defined in terms of language competence in his work. This list of constructs shared similar listening abilities to be assessed as those identified in terms of communicative aspects. The difference is that the competence constructs were viewed from competency aspects. His framework consists of the following elements.

- knowledge of the sound system: It includes those relevant aspects of grammatical knowledge, namely phonology, stress and intonation.
- understanding local linguistic meanings: It includes the whole grammatical knowledge, not only phonology, stress and intonation, but also vocabulary and syntax, as well as the ability to use that knowledge automatically in real time.

- understanding full linguistic meanings: It includes grammatical knowledge plus discourse knowledge, and would require understanding longer texts.
- understanding inferred meanings: It includes grammatical knowledge, discourse knowledge, and pragmatic knowledge – that is understanding inferred meanings and unstated implications.
- communicative listening ability: It includes grammatical knowledge, discourse knowledge, pragmatic knowledge and sociolinguistic knowledge; this would be communicative language ability.

To compare the constructs of the listening skills provided by Aitken (1978) and Weir (1993), Table 2.3 shows the constructs of the listening skills and sub-skills listed by two researchers. These constructs are taxonomy of 'communicative' listening sub-skills. Weir's list tended to be more complete than Aitken's. However, both insisted not to take their framework as a complete list of all listening sub-skills (Buck, 2001). In other words, the listening constructs to be tested may involve other sub-skills, which were not discussed in their work.

Table 2.3

A Comparison of the Listening Constructs Proposed by Aitken (1978) and Weir (1993)

Weir (1993: 98-99)	Aitken (1978) in Buck (2001: 54)
Direct meaning comprehension: - Listening for gist - Listening for main idea(s) or important information, including tracing the development of an argument, distinguishing the main idea(s) from supporting detail, differentiating statement from example, differentiating a proposition from its argument, distinguishing fact from opinion when clearly marked. - Listening for specifics, involving recall of important details. - Determining speaker's attitude/intentions towards listener/topic (persuasion/explanation) where obvious from the text	- Identifying the speaker's purpose - Recognizing the speaker's attitude to the listener and the subject of their discussion
Inferred meaning comprehension: - making inferences and deductions: evaluating content in terms of information clearly available from the text - relating utterances to the social and situational context in which they are made - recognizing the communicative function of utterances - deducing meaning of unfamiliar lexical items from context	- drawing correct conclusions and valid inferences about the social situation, the speaker's intent of the general context - being able to guess the meanings of unfamiliar or unclear words from their context

Table 2.3 (Continued)

Comparison of the Listening Constructs Proposed by Aitken (1978) and Weir (1993)

Weir (1993: 98-99)	Aitken (1978) in Buck (2001: 54)
Contributory meaning comprehension (microlinguistic):	
- understanding phonological features (stress, intonation, etc.)	- understanding the flow of stressed and unstressed sounds, and intonation cues and other cues of oral punctuation
- understanding concepts (grammatical notions) such as comparison, cause, result, degree, purpose	- understanding the syntactic patterns, the morphological forms, which are characteristic of spoken language, and following the discourse patterns of spoken language
- understanding discourse markers	
- understanding syntactic structure of the sentence and clause, e.g. elements of clause structure, noun and verb modification, negation	
- understanding grammatical cohesion, particularly reference	
- understanding lexical cohesion through lexical set membership and collocation	
- understanding lexis	- understanding the vocabulary

Table 2.3 (Continued)

Comparison of the Listening Constructs Proposed by Aitken (1978) and Weir (1993)

Weir (1993: 98-99)	Aitken (1978) in Buck (2001: 54)
Listening and writing (note taking from lectures, telephone conversations, etc.): - ability to extract salient points to summarize the whole text, reducing what is heard to an outline of the main points and important details - ability to extract selectively relevant key points from a text on a specific idea or topic, especially involving the coordination of related information	

In Table 2.3, a comparison of the listening constructs proposed by Weir (1993) and Aitken (1978), cited in Buck (2001), is presented to make a conclusion of which listening constructs are required to be tested. It consists of four main categories which are direct meaning comprehension, inferred meaning comprehension, contributory meaning comprehension, and listening and writing (note taking). All constructs were explained in the earlier part. In the first category, 'direct meaning comprehension', Weir listed four constructs. They include listening for gist or listening for general ideas, listening for main ideas or important details, listening for specific details, and determining the speakers' attitudes or intentions. The last construct agrees with what Aitken proposed as test takers' ability to identify the speaker's purpose and recognize the speaker's attitudes toward the topic.

The second category, 'inferred meaning comprehension', is composed of four sub-constructs under Weir's (1993) framework. They include making inferences and deductions by using the context, relating utterances to the social and situational context, recognizing the communicative function of utterances and deducing meaning of unfamiliar lexical items from context. In Aitken (1978), cited in Buck (2001), the constructs proposed are similar. Aitken's 'drawing correct conclusions and inferences' also includes the evaluation of test takers' ability to relate the utterances to the social

and situational contexts, whereas Weir (1993) separated these constructs into four sub-categories, which are similar to Aitken's (1978) proposal. Moreover, both researchers integrated the construct that measures test takers' ability to guess the meaning from contexts.

The third category, 'contributory meaning comprehension', proposed by Weir (1993) involves seven sub-constructs: understanding phonological features, understanding grammatical concepts, understanding discourse markers, understanding syntactic structure of sentences and clauses, understanding grammatical cohesion, understanding lexical cohesion, and understanding lexis. Like Weir, Aitken (1978), cited in Buck (2001), provided comparable listening constructs which are those measuring test takers' ability to understand the stressed and unstressed sounds, intonation, and other phonological features, to understand the syntactic structures of the spoken texts, and to understand the vocabulary.

The fourth category, 'listening and writing', proposed by Weir (1993) was not integrated in Aitken's (1978) work. This is because the constructs in this category are optional in language listening tests. And, there are many of them, including the TOEIC, which do not measure test takers' ability to listen, take notes and make use of their notes.

In the next part, the listening claimed to be assessed by various standardized tests would be analyzed and categorized according to Weir's (1993) framework. Examples of each sub-skill are given below.

Listening Constructs of CULI Test PIC

Following are the listening constructs claimed to be tested by the CULI Test PIC. The information was provided by Chulalongkorn University Language Institute.

1. Direct meaning comprehension

- listening for gist: identifying the paraphrase or restatement of what the speakers have said

- listening for main idea(s): identifying topic, main idea, title

- listening for specifics: identifying specific details

- determining a speaker's attitude or intentions: identifying tone, attitude of the speaker, intended audience, purposes

2. Inferred meaning comprehension

- making inferences and deductions: drawing conclusions, making inferences, making predictions (e.g. setting, role, status, seniority, emotion, mood, etc.), identifying sources of listening (Where would you hear this announcement?)

- deducing meaning of unfamiliar lexical items from context

3. Contributory meaning comprehension (microlinguistic)

- understanding phonological features: minimal pairs, homophones

- understanding lexis: idioms

- understanding grammatical notions: causative verbs

Listening Constructs of TOEFL

The TOEFL includes the following listening constructs (Gear and Gear, 2002).

1. Direct meaning comprehension

- listening for gist: identifying the restatement of what the speakers have said

- listening for main idea(s): identifying topics, suggestions, planning, problems, etc.

- listening for specifics: identifying specific details

- determining a speaker's attitude or intentions: identifying attitude, opinions, purpose of the speaker

2. Inferred meaning comprehension

- making inferences and deductions: drawing conclusions, making inferences, making predictions (e.g. future actions), making assumptions

3. Contributory meaning comprehension (microlinguistic)

- understanding phonological features: minimal pairs, homophones

- understanding grammatical notions: comparison, cause, effect, conditions

- understanding lexis: idiomatic expressions and phrasal verbs

Listening Constructs of IELTS

Another well-known and widely used standardized test, the IELTS, includes the following constructs in the listening part (O'Sullivan and Thurlow, 2002).

1. **Direct meaning comprehension**

- listening for gist
- listening for main idea(s)
- listening for specifics: identifying times, dates, names, keywords
- determining a speaker's attitude or intentions

2. **Inferred meaning comprehension**

- making inferences and deductions: making predictions, making inferences about where the speakers are, what they are speaking about, and who they are
- deducing meaning of unfamiliar lexical items from context

3. **Contributory meaning comprehension (microlinguistic)**

- understanding phonological features: intonation, stress
- understanding grammatical notions: comparison, cause, effect, process, classification, argument
- understanding lexis
- understanding discourse markers: but, and, although, etc.

4. **Listening and writing (note taking)**

- ability to extract salient points to summarize the whole text
- ability to extract selectively relevant key points from a text

Listening Constructs of TOEIC

The TOEIC, which aims at measuring test taker's English proficiency concerning the area of language use at work, integrates the following listening constructs (Gilfert, 1996).

1. Direct meaning comprehension

- listening for gist
- listening for main idea(s): identifying purpose of the message
- listening for specifics: identifying problem, topic
- determining a speaker's attitude or intentions

2. Inferred meaning comprehension

- making inferences and deductions: making predictions (future actions), making inferences about where the speakers are, what they are speaking about, and who they are

3. Contributory meaning comprehension (microlinguistic)

- understanding phonological features: homophones
- understanding grammatical notions: negatives, question tags, etc.
- understanding lexis

The analysis of the standardized proficiency tests showed similar listening constructs being tested; therefore, this supported the framework proposed by Weir (1993), and confirmed that these are the listening skills that must be tested in a proficiency listening test.

2.3.4 Creating listening tests

After the construct of the test is clearly defined, listening tasks can be designed. Littlewood (1981) categorized listening physical tasks into five groups, namely identification and selection, sequencing, locating, drawing conclusions as well as performing other actions. The first type of task is listening to a description and selecting the picture to which it refers. For sequencing tasks, learners have to put their answers in order; for example, they hear a story and have to put the pictures relating to the story in chronological order. The third type of tasks is locating which can be best explained by the task concerning giving and following directions. The listener has to find the place on a map. Examples of drawing and constructing tasks are like drawing a picture based on the description heard or completing an outline of a spoken text. Last but not least,

regarding performing other actions, learners are asked to perform an action after hearing an input.

In language assessment, samples of communicative language ability and identification of real-life conditions or situations need to be replicated so that teachers can make valid statements of students' ability. In testing listening skills, characteristics of the spoken language must be incorporated. Heaton (1990) suggested the following characteristics of spoken language that should appear in an assessment.

Characteristics of the tasks designed must be considered so that they really match the listening skill. This means characteristics of spoken language must be replicated in a good listening test. Heaton (1990) provided a list of spoken language's features.

1. Repeating information or redundancy

It is very noticeable that spoken language contains a lot of redundant parts and meaning is usually reinforced or repeated in several ways, e.g. 'Would you care for a game of tennis?' 'Tennis?'

2. Pausing

Spoken language is usually unplanned speech; therefore, it is filled with pauses and fillersm e.g. 'Er...what do you...em...would you come for a game of tennis?'

3. Differences between speaking and writing

Speaking and writing are different modes of communication. Writing requires planning and revising, whereas speaking does not. Writing contains structures that are complicated, whereas speaking does not.

4. Things that we remember

The message that is stored in our brain or that we remember is general meaning rather than actual words found in the spoken text.

5. The importance of context

Sometimes the same sentence can be interpreted in many ways based on the context, e.g. 'Linda talked a lot.' This sentence can be criticism, a favorable comment, etc. depending on the situation.

6. Using recorded material

Spoken language that is recorded for a listening test is usually recorded, which makes the listening test deprived of other non-verbal languages such as gestures, facial expressions, etc.

From the list of characteristics of general spoken language, the following implications for listening tests must be taken into account by teachers.

1. The speaker in the listening test must talk, not read.

- Rewrite most complex and long sentences.

- Rewrite the talk and repeat the important points.

- Pause slightly longer at the end of clauses and sentences rather than reading texts very slowly.

Mead and Rubin (1985) agreed that the listening stimuli should represent typical oral language. Simply reading written material can cause problems.

- When writing questions, keep in mind that students do not see the written form.

2. There is both an advantage and a disadvantage when using recordings.

- The benefit is that it is more reliable. No matter how many times you play the recording, the same speeches are repeated.

- The drawback is that it is more difficult than real life conversations given where interlocutors can also rely on a non-verbal input.

3. Length of talks does matter.

- A long talk is more demanding.

- Instructions must be kept simple.

The characteristics of spoken language lead to implications for listening tests that some teachers might forget. Firstly, the assessment must avoid reading aloud along written texts that lack redundant features, which are considered important for students' comprehension. Secondly, simple language must be used in instructions and questions to reduce students' anxiety which can negatively affect students' performance.

Moreover, the fact that listening performance is affected by motivation and memory, it is suggested that the passages selected be interesting and relatively short. In addition, topics should not be bias in terms of sex, geographic, socioeconomic, or racial/ethnic background. And the environment where the assessment and evaluation takes place should be free of external distractions. In other words, the sound quality must be excellent, the questions are printed clearly, and the speed in delivering messages must be right, and so forth (Mead and Rubin, 1985).

2.3.4.1 Communicative language testing

According to Ellis (2003), the concerns towards communicative language testing were reliability and validity. An attempt to assess language ability in general was the Communicative Use of English as a Foreign Language Test (CUEFL) developed by Royal Society of Arts (RSA) in the U.K. Marks were given to overall language ability, rather than isolated linguistic items.

It is argued that communicative language testing found in the past failed to address reliability and construct validity of tests. Bachman (1990) proposed a modular model of language knowledge consisting of organizational knowledge and pragmatic knowledge. The former refers to grammatical and textual knowledge, while the latter relies on functional and sociolinguistic knowledge. In addition, he also included strategic competence so that the way these knowledge sources were integrated into actual performance could be explained.

Fulcher (2000), cited in Ellis (2003) defined a communicative test in three primary aspects: performance, authenticity and real-life outcomes. This clearly shows that

communicative language testing constitutes a form of task-based assessment. It gives value to real-world tasks, although it is still doubtful if tests tasks really replicate real-world ones.

2.3.4.2 Definitions of communicative tasks

Task-based instruction has long been established as a communicative instructional approach. Various definitions have been proposed to limit the scope of tasks. Long (1985: 89) cited in Celce-Murcia (2001) defined a task as:

> "A piece of work undertaken for oneself or others, freely or for some reward...examples...include painting a fence, dressing a child, buying a pair of shoes... 'task' is meant the hundred and one things people do in everyday life, at work, at play, and in between."

Crookes (1986) refers to a task as a purposeful activity carried out in classroom or at work. In Ellis (2003: 4), the definitions of tasks given by researchers were spelled out as follows:

Breen (1989) defined a task as a structured plan aiming to provide opportunities of the refinement of knowledge and capabilities required in communication. Breen specifically states that a 'task' can be 'a brief practice exercise' or 'a more complex workplan' that requires spontaneous communication of meaning.

Long (1985) provided a definition of a task as work done for oneself or for others, freely or for some reward. He gave various examples of tasks, e.g. painting a fence, dressing a child, filling out a form, buying a pair of shoes, making an airline reservation, borrowing a library book, taking a driving test, typing a letter, weighing a patient, sorting letters, taking a hotel reservation, writing a cheque, finding a street destination, helping someone cross a road, etc. For Long (1985), tasks are things people do in everyday life. They are engaged with all of our activities.

Prabhu (1987) said a task is an activity involving some processes of thought about the information given by teachers before learners achieve their goal. A task allows language teachers to control and regulate that process.

Murphy (1993) defined tasks as things we do in our daily lives, and, in order to complete these tasks, we must learn how to do them. It is related to reality, goals and routines.

Bachman and Palmer (1996) defined a task as an activity that involves individuals using a language for the purpose of achieving a particular goal or objective in a particular situation.

The definition of tasks proposed by Skehan (1998), who has been actively involved in the area of task-based instruction, seems to be the one quoted in many textbooks. Skehan's definition of a task is that it is an activity that primarily focuses on meaning, involves communication (e.g. to solve problems), is related to real world activities, has a goal to work towards and is assessed in terms of outcome.

From the definitions of everyday tasks is derived the concept of pedagogical tasks. Richards, Platt, and Weber (1985: 289) refer to a task as:

> "An activity or action, which is carried out as the result of processing or understanding language. For example, drawing a map while listening to an instruction and performing a command... A task usually requires the teacher to specify what will be regarded as successful completion of the task."

Lee (2000) said a task is a classroom activity or exercise that has an objective obtainable only by the interaction among participants, a mechanism for structuring and sequencing interaction, and a focus on meaning exchange. Also, it is a language learning endeavor that requires learners to comprehend, manipulate, an/or produce the target language as they perform some set of work plans.

Based on the definitions above, a pedagogical task should be defined as a real world related activity performed in a language classroom to practice the target language. It is goal-oriented and focuses on communicating meaning. Therefore, the concept of a language task is totally opposite that of traditional paper and pencil exercises. Murphy (1993) defined two scopes of language activities as mechanical exercises and communicative tasks. The first is related to the structuralist approach whereas the latter is associated with such tasks as simulation, information gap or jigsaw tasks.

What do tasks used in language testing and evaluation refer to? According to Ellis (2003), task-based language testing can be defined based on two aspects. Firstly, it complies with the characteristics of direct system-referenced tests. That means that language proficiency is assessed as a system and is construct-oriented whereas the contextualized sample of the testee's use of language is selected. Secondly, it can be viewed as part of direct performance-referenced tests, which also means the contextualized sample of the target language use is identified. However, opposite the view towards system-referenced tests, it is more content-oriented and seeks to provide information about the ability to use the language in specific contexts, rather than the knowledge of language as a system.

2.3.4.3 Types and features of tasks

Since it is believed that tasks can serve as tools to set an environment that promotes language acquisition, various kinds of tasks are designed and used by teachers to generate desirable outcomes. Such tasks as picture difference, role-play, story telling and scavenger hunt are very common and usually found in language classrooms.

Crookes and Gass (1993) reported two categories of tasks: one-way tasks and two-way tasks. These categories are simply based on the direction of information. If one student produces information and others listen, such as picture drawing where one student gives instructions and the other draws, that task is called one-way. However, if

the information flows from both of the interlocutors, the task is called two-way. Information gap tasks are perfect two-way tasks where both speakers have information and they take turns exchanging information. Duff (1986), cited in Crookes and Gass (1993) separates "divergent from convergent tasks" in her study. The first aims at having students argue their positions by giving logical arguments, whereas the latter's goal is towards mutually acceptable solutions or consensus. To distinguish tasks, Duff focuses on interactional direction and the nature of the content.

Since the term "tasks" is very broad, some researchers attempt to make a clear distinction by categorizing tasks into target tasks and pedagogical tasks. The first refers to tasks that students must accomplish outside classroom while the latter, according to Brown (2001), involves some techniques designed to instruct and train students to be able to complete the target task. For example, students would like to learn what to say when they order food in a restaurant. Their pedagogic task will be designed so that they can practice the skill required to accomplish the target task. Pedagogic tasks' goals are beyond the classroom level. They aim at the target tasks that students must be able to perform.

To prepare for task-based instruction, teachers must firstly understand the features of tasks so that they can expect whether the tasks to be used will formulate what kind of product, which kinds of operations are required to accomplish the tasks, and whether there are any resources available to the students to promote the expected production (Nunan,1988).

Robinson (2000), cited in Celce-Murcia (2001) proposes three aspects of tasks: complexity, conditions and difficulty. Robinson treats these aspects as separate units. Task complexity is relevant to the factors affecting learners' cognition that will lead to language accuracy, fluency and complexity. For example, if students are allowed to plan before performing a task, the production will be more accurate, fluent and complex. Sometimes, they occur at the same time, while sometimes, there seems to be a trade-off. Task conditions affect the amount of production, interaction and feedback from learners.

According to Robinson (2000), tasks which influence "participation variables" are open and closed tasks, one-way and two-way tasks, and convergent and divergent tasks. Tasks possessing these features have an influence on learners' interaction. Task difficulty deals with learners' ability to perform an assigned task. Factors that are bound with task difficulty are, for example, motivation, anxiety, confidence, aptitude, level of attained proficiency in the L2, and intelligence.

2.3.4.4 Listening tasks

Ellis (2003) mentioned two kinds of listening tasks. To begin with, non-reciprocal tasks best describe what listening tasks are like since they are passive in a way that rarely do listeners interact to the text they hear. On the other hand, reciprocal tasks require both interlocutors' negotiations such as in information gap activities. Therefore, research in listening tasks usually focuses on the non-reciprocal paradigm while there are two types of these particular tasks proposed: listen-and-do tasks and academic listening tasks.

The first kind of task, listen-and-do, relies on research in the task-based teaching area. When performing this particular task, students have to listen and show their understanding by accomplishing the task, e.g. listening and following directions.

Academic listening tasks require learners to listen to lectures and show their understanding in certain ways such as by taking notes. The format is relatively universal. The topics are on academic issues. Some may argue that taking notes should not be considered as tasks; however, it meets the criteria of task features such as focusing primarily on meaning to reach the goal, using linguistic knowledge to process the input and to produce the notes. It truly engages a lot of cognitive process since students have to identify the structure of the content and select key points to be reported.

Based on research studies in the past, this task is effective in providing meaning negotiation among adolescent and adult learners. Negotiation of meaning enhances comprehension. However, for acquisition, it was found that the process is different.

The research on academic listening tasks concentrates on the discourse features of the lectures that serve as input. Such factors as lecturing styles (non-active and active), the discourse structure, interpersonal features and lexico-grammatical features are of researcher interest. Until now, no research has been done to explore how academic listening tasks affect language acquisition. However, it is clear that academic listening tasks create a 'need' for learners to process words in the input. So far, it shows that academic listening tasks serve as a means by which learners can expect their vocabulary.

2.3.4.5 Tasks in language assessment

How does the task-based approach enter the realm of assessment? As language instruction currently points towards communicative goals, the focus of language teaching and learning has shifted from discrete components of language to the emphasis on language use. In the past, listening assessment aimed at testing students' ability to distinguish sounds, recognize stress and intonation pattern and record their answers through multiple choice. This poses questions as to whether the ability to discriminate those discrete points can describe students' ability to comprehend verbal input. At present, communication of meaning has received more attention. Language tests, then, concentrate more on language use such as summarizing what students have heard, following directions, and filling out forms (Weir, 1993). This results in the increasing use of tasks as means to assess students' ability to use language.

This is influenced by the communicative competence theory that was pushed in the 1970s. The use of 'task' has been increasingly promoted. Language does not consist of the production of separate sentences, but the use of sentences to create discourse (Chalhoub-Deville, 2001). Although linguists and researchers have proposed various components of communicative competence, they mostly share similarities. For example, Canale and Swain (1980) proposed that communicative competence embraces the competence in the area of grammar, sociolinguistics and strategy in using language to communicate. Similarly, Bachman and Palmer (1987) integrated grammatical and

sociolinguistic competence. However, they include pragmatic competence while eliminating strategic competence. Despite slightly different views towards communicative competence, it can be concluded that communicative competence is the combination of organizational competence (both grammatical and discourse), pragmatic competence (both functional and sociolinguistic), strategic competence, and psychomotor skills. Communicative competence can be achieved by paying attention to language use, fluency and accuracy, authentic language and context, and a connection between classroom practices and those in the real world (Brown, 2001). If how we assess students must reflect how we teach them, tests such as multiple-choice tests, definitely cannot serve this purpose. This is perhaps why task-based assessment is turned to.

Although people argue that there is no assessment that purely reflects real-life language use, and genuineness of text may not always be feasible in tests, authenticity of tasks in language testing has gained more attention from teachers and test developers. They are more careful in selecting and developing tasks to ensure that features of real life language use, as many as possible, are incorporated in the test as well as to make sure that the conditions in tests resemble real-life situations (Weir, 1993).

2.4 Issues concerning the integration of tasks in language tests

This part introduces advantages and concerns over the use of tasks in assessment.

2.4.1 Benefits and Limitations of tasks

Since this is the era of communicative language teaching, the orientation towards communicative goals is emphasized more than the production of correct grammatical forms as was focused on in the past. And, if we explore the features of tasks, we will find that tasks are perfectly matched with the communicative teaching approach. The advantages gained from the use of tasks in classrooms basically derive from their characteristics. Firstly, because pedagogic tasks are based on target tasks, as suggested by Long (1997), pedagogic tasks can be designed to prepare students to

perform actual tasks in the real world. For example, teachers can have students practice job interviews in class to prepare them for the real job interviews that they will face in the future. It can be seen from this example that tasks can benefit students more than just giving them a chance to practice the language in the desirable context, but they can also provide students with the whole picture or setting of what the real life target task will be like. This can help reduce students' anxiety when they perform the target task as well.

Moreover, tasks support the student-centered learning style. Chalhoub-Deville (2001) said tasks can strengthen the idea since they present test-takers with novel situations embedded with context descriptions or visuals. They, then, can activate and rely on their schemata to attain the communicative goals. When using tasks, students usually work in groups or pairs. Teachers become facilitators trying to generate discussions among students. While students are discussing or doing the task by using the target language, they have to use several strategies to negotiate for meaning and to help them overcome the communication breakdown. From this point of view, if we consider the theory of second language acquisition, it is undeniable that tasks set an appropriate environment that can promote language acquisition. According to Long (1997), as students negotiate for meaning to complete the task, they learn new knowledge, perhaps through feedback or recasts. Sometimes students correct one another. And their peers do repeat the correct form. It means that they pay attention to the language, and attention is one of the main factors that catalysts language acquisition.

In addition, since tasks emphasize authentic language, Guariento and Morley (2001) found that authentic text can maintain or even promote motivation in learning or increase students' attention. Students feel that the language they are exposed to is similar to that used in real life.

Finally, regarding evaluation, a paper and pencil examination might not be enough to measure students' ability to complete tasks. There must be a more suitable way to replace or use with this traditional examination. Assessing their performance by giving them tasks with the same goals can serve as an alternative. For example, if

teachers would like to see if students can give directions and they have already practiced that in class, teachers can use different maps when evaluating them. In this case, students will still have to use the structures or forms required for giving directions, but in a different context. Teachers can also observe their performance and their participation when performing tasks. This type of assessment can be used to balance the proportion of students' marks as well as to measure students' ability to use the language in a communicative way.

Apart from the benefits gained from using tasks, there are several limitations discovered when tasks are used. The development of good tasks is time consuming and requires a lot of preparation as can be seen from the steps mentioned earlier (Long, 1997). There are also some complications in terms of preparation and classroom management.

Despite the belief that tasks drive language acquisition, we cannot be certain that the positive environment for language acquisition that tasks generate will always lead to students' interlanguage construction. This is because it deals with internal mechanism that we cannot observe. Students may learn to produce the right form when doing the task at the moment, but forget it later. Although it is a common process of language acquisition, teachers should be aware of it and perhaps find a way to improve the efficiency of tasks or implement post-tasks to strengthen the language acquisition process.

One of the most significant aspects that teachers cannot overlook is that they cannot only place themselves as the center of educational management. There are many factors relating to learners that can fail the best teaching method. Since learning and teaching include two parties, teachers and learners, learners' styles should also be taken into account. For example, adult learners might not be familiar with practicing the language with tasks and feel uncomfortable in class. Some learners do not like group work or pair work, but prefer individual assignments. Frequently, when students of the same L1 background first meet, they feel awkward to use the second language with their

peers. In this case, tasks can be delayed instead of being given at the beginning of the class.

2.4.2 Problems found in task-based assessment

Ellis (2003) pointed out the following problems in task-based assessment. The first problem is representativeness of tasks that are integrated into a test. Wigglesworth (2001) suggested that the test should contain sample conditions of what test-takers are going to face in the target language use domain. Brown (1997) also said that to ensure the representativeness, the type of performance to be measured must be overtly stated. The second factor is concerned with authenticity of tasks. Bachman (1990) insisted that to achieve authenticity, both situational and interactional authenticity must be met. The third problem is the issue concerning generalizability of tasks. This notion refers to the extent to which the test tasks can be found in the real world, or how they can be generalized to the real world tasks. Another factor that should be taken into consideration involves inseparability of subject content knowledge from language knowledge. It is believed that some subject content cannot be separated from the knowledge of language. Background knowledge and language knowledge are viewed as 'inextricably intertwined', especially in ESP. Lastly, the reliability of tests must be considered. Several sources that can cause unreliability of the task-based tests are the personal dispositions of the candidates shown during the testing process, e.g. anxiety, administrative conditions and inconsistency in scoring procedures.

2.5 Computer-based Assessment

The role of the computer in language assessment has been established due to the benefits it offers. However, there are still some arguments against the use of the computer. This section, then, reviews the roles of the computer in language assessment, the advantages and the disadvantages of the computer as a tool to assess language abilities, and other factors involving creating and implementing a computer-based test.

2.5.1 The roles of computer in assessment

Alessi and Trollip (1991) proposed two major ways to integrate computers in the testing process. The first is to use the computer as a tool to construct or score tests. For example, the computer is used to create a shared pool of questions. Teachers write questions and store them in pools or item banks. These item banks are accessible when needed. Sometimes, not only are the item banks shared among teachers in the same institution, but they are also open to teachers from various places. The second role of the computer is to administer tests. That is, the entire test is automated and students have to take the test using a computer, instead of using paper and pencils. More elaborate roles of computers are described by Stephens, Bull and Wade (1998). The activities in which the computer is used in the assessment are to deliver, mark and analyze examinations; to record, analyze and report on achievement; to collate and analyze data obtained from optical mark readers; and to collate, analyze and transfer assessment information through networks.

2.5.2 Advantages and disadvantages of the use of computers in assessment

It is undeniable that the computer offers numerous advantages over the traditional method of assessment. Firstly, according to Alessi and Trollip (1991), it serves as a good assistant for teachers. With its capability to carry a large pool of questions that teachers can share and randomly distribute to test takers, teachers do not have to write a lot of items themselves, but share them with their colleagues. It also helps teachers save time in scoring students' examinations, especially a large number of them. In addition, teachers can store students' answers as a record for future improvement of the test. For students or test takers, carefully designed computer-based tests are individualized. By means of the computer, a test can be tailored to the ability levels of each test taker.

In addition, with advanced technologies, computer-based tests can be more authentic and more interactive than a paper and pencil test (Bachman, 2000). Instead of interacting with students through multiple choice (MC), computer-based tests can make interactions through other forms of responses more attractive to test takers, e.g. clicking

the picture described in the lecture. Furthermore, students are allowed to take a test when they are ready, rather than at a fixed time. The same test can be constructed differently for each individual. Some computer-based tests, such as TOEFL, give immediate results to the students. This helps them decide at an earlier stage which step they will move towards in the future; for example, they may decide to take the test again next month and start studying for the next test right away.

What teachers are concerned with and have observed in various studies are students' attitudes towards the use of the computer in language assessment. A study conducted by Bocij and Greasley (1996) shows test takers' positive attitudes towards computer-based assessment. In his study, the scores students obtained from a computer-based test were higher than those gained from a conventional exam. This may have resulted from less time they spent on the test. If the performance is related to the time taken to complete the assessment, then the computer can help students spend less time to answer questions leading to more time to work on the test.

Moreover, it is interesting to see that students feel more relaxed when sitting in a laboratory as opposed to an examination hall. They also feel that the computer-based test is quicker to work through. The confirm function that the computer offers before students can move to the next question forces them to reread the questions and be more careful. However, teachers must explore whether computer-based testing creates positive atmospheres and testing conditions.

Despite the impressive qualities that the computer possesses, teachers and administrators still have to consider its limitations. According to Alessi and Trollip (1991), being able to access the item banks can pose certain problems. For example, the items drawn from the item banks may be abused and misused by teachers. When using the questions from the pools, teachers must make sure that the questions reflect the objectives of their instruction. The quality of questions is another problem. New items added to the pools may not be tested. And, the permission to access the pools from anywhere can cause security problems. Another disadvantage of computer-based tests

is the format of the test is restricted to multiple choice, short answers and matching due to the difficulty in grading answers to such items as essays by the computer.

Bachman (2000) raised a question whether the constructs of a test that requires reading a text from a test booklet and that which is presented on a screen are similar. In addition, some external difficulties like power failure or insufficient terminals can lead to severe logistical difficulties. Back-up procedures should be planned in advance. Last but not least, there is the major concern of students' anxiety and frustration. This problem can be solved by introducing the use of computers in class through certain activities, e.g. collaborative web-based projects (Lewis and Atzert, 2000) so that students can be more familiar with the computer. The software used to administer the test which is user-friendly can help reduce the anxiety. However, anxiety may result from the test format as well. As cited in Shabaan (2001), Cohen (1984) reported that open-ended questions are preferred to multiple choice questions, whereas the cloze test is perceived as the most difficult. However, more recent research must be conducted to see if today's learners' views have changed.

Therefore, interface development and design are crucial in computer-based language testing since they can threaten construct validity resulting in misinterpretation of students' language ability. "The user interface is the elements of a software program that a user sees and interacts with. The more intuitive the computer test software is, the less the examinee needs to attend to it. A good interface is clear and consistent and should be based on sound software design principles in order to support the overall goals of the program." (Parshall et al, 2001: 5) Fulcher (2003) emphasized good interface design the primary goal of which is to reduce to a minimum construct-irrelevant variance in testing. The ultimate aim is to ensure that the mode of delivery does not contaminate the scores that means there is the least error in measuring language ability.

2.5.3 Three phases of designing a computer-based test

2.5.3.1 Phase 1: planning and initial design

This phase initially involves such work as establishing the design team, identifying test takers, stating test purposes, and describing test constructs. After that, content is defined. Item prototypes and scoring systems are then designed based on the content. Later, an interface prototype, a scaled-down preliminary version of a test interface consisting of samples of content of item types to be included in the test, is designed. This prototype must not negatively affect the validity of the test. In this phase, all components relating to test interface design, e.g. navigation, terminology used in instruction, page layout, text, color, icons, etc. are taken into consideration.

2.5.3.2 Phase 2: usability testing

In this phase, test developers trial the test with a small number of subjects to search for problems and solutions. This group of subjects must represent the target group of learners who are going to take the test, have a range of experience of computer interfaces, have a variety of expectations of a CBT and possess different language ability. Fulcher (2003) suggested one or two designers observe test takers and make notes on their performance. A debriefing interview should be conducted at the end. Also, the "think aloud" technique is often used while the test takers are taking the test. During that stage, the observers record the protocols, which will be used in the analysis. Interruption when students are facing problems in navigating through the test is acceptable. After the observation, structured interviews are recommended so that the observers can elicit problems and find logical solutions.

2.5.3.3 Phase 3: field testing and fine tuning

During this phase, a large sample of the test-taking population is required to participate. There should not be any more major changes. According to Fulcher (2003: 403), "Field testing is primarily required for scaling the test and ensuring that the logistics of data collection, submission, scoring, distribution and retrieval, and feedback work as planned....It does, however, provide an opportunity to test for variation in the appearance of the interface across sites, machines and platforms."

2.5.4 Implementation of computer-based assessment

Few research studies reveal a model to implement computer-based assessment. An attempt was made by Dunkel (1991), who tried to create a computer-based listening test that could be used with micro computer hardware as well as contain the taxonomy of conversational and academic listening skills described by Richards (1985).

According to past literature, three levels of implementation are found. At the level of administration, examples are shown in Stephens et al (1998). University of Luton uses a top down approach beginning with Faculties and Departments. Then, a computer-assisted assessment unit is established to provide support before the academic staff is encouraged to participate. The next step involves the implementation of a student-centered curriculum with computer-assisted assessment as a key component. At this point, students in class are made familiar with the use of computer assisted assessment. On the other hand, Loughborough University adopts a bottom-up approach starting from the academic staff. However, no matter which approach the universities use, they are all engaged in careful planning before integrating the new kind of assessment into the curriculum. For instance, the computer assisted unit is established and students are introduced to the assessment.

For overall test development, Alessi and Trollip (1991) suggested two major phases: determining the characteristics of the test and administering the test. First of all, the purpose of the test must be determined. For example, is it criterion-referenced or norm-referenced testing? After that, the objectives of the test need to be set up. They recommended that the instructional objectives be listed to guarantee all topics are covered. If the amount of content outweighs the amount of time, it is better to increase the number of tests rather than reducing the content. Bachman (2003) emphasized that investigation of the TLU domain and identification of tasks will help reflect test takers' true abilities. Other aspects needing to be considered are length, feedback, timing and presentation of results. For the implementation phase or when designing a computer-based test, the following factors are required. Firstly, access to required information is needed. Students may need to access the instructions and teachers may need to

access to the test results. Secondly, users should have great control over the program. They are able to decide what to do next, rather than following the computer's instruction. For example, on a TOEFL test, test takers can take time answering each question and decide when to move on. They can take more time on one item and less time on another. Thirdly, safety barriers and safety nets must be installed to prevent accidental termination of the assessment. Lastly, the role of students must be placed in all stages.

2.5.5 Factors affecting students' performance

What has been focused on by researchers more and more are the factors affecting students' performance when they are taking a test, either a computer-based or a conventional test. Bachman (2000) proposed three broad areas, namely characteristics of the testing procedure, strategies and processes used by test takers when answering questions, and test takers' characteristics.

There are significant relationships between item difficulty and characteristics of the test items. If the level of vocabulary and syntax is complex or if the cognitive demand is high, students will find it difficult to find the right answer. Task types also affect performance of test takers. The study of Yi'an (1998) revealed interesting results about the performance of students in a multiple-choice task of a listening test. The MC format favors the advanced test takers who possess more competent linguistic processing. The choices given facilitate the advanced listeners, whereas they do not help less competent learners but lead them to wrong guesses. However, options can lead to failure in answering a question as well if students cannot interpret the meaning of two options. Even though this study conveys such informative data, it is viewed from the perception of a researcher through the quantitative method. The number of correct answers may not tell much about students if their attitudes are ignored because there are many other factors that can affect their scores. Also, more research on how different kinds of task types affect students' performance is needed so that the test format can be improved.

Visuals are also found to be a factor affecting students' performance. Duzer (1997) suggested that visual support such as video, pictures, diagrams, gestures, facial expressions, and body language can enhance comprehension. However, she emphasized that this positive environment will happen only when listeners are able to correctly interpret them. According to Parshall, et al (2001), the visual style should suit the target audience and the subject matter of the program. A simple, uncomplicated style is most acceptable for computerized testing. Clear, legible text and few screen windows opening at one time is best, especially for inexperienced users. Soft colors and obvious contrast between the front and the background help reduce fatigue during the exam.

Ginther (2002) reported how context and content visuals play a role in listening computer-based assessment. The study was done with the TOEFL examination in which both types of contexts are present. The context visuals refer to those that provide information about the context in which the verbal exchanges occur, e.g. visuals showing where the conversation takes place or which person is talking, whereas the content visuals are those that relate to the content of the utterances. The result shows greater performance with the presence of the visuals, especially the content visuals. The other study cited by Bachman (2003) in his presentation showed level of proficiency interacts with how they find visuals helpful and which strategies they use in responding to different types of visual input (Kim, Kim and Shin, 2001).

Sherman (1997) studied the effect of question preview on listening comprehension tests. The research focuses on four types of question preview to see the most appropriate time to show questions to test takers. Version A (Questions Before) involves previewing ten questions that are answered during and after two hearings of the text. In Version B (Questions After), the test takers listen to the text twice and questions are given later. Therefore, they do not see the questions until they finish listening. Version C (Sandwich), they listen once and read the question before listening again. And in Version D (Free), they hear the text twice, and then write down what they can

remember about the verbal input. The study showed that Version C positively affects students' performance the most whereas Version D causes the most problems.

Secondly, strategies and processes used by test takers when answering questions seem to be related to the cognitive processing or communicative competence that each test taker possesses. Different types of tests can lead to different strategies used. The strategies employed by individuals will affect how well they perform.

Lastly, other factors are related to learners' characteristics, e.g. gender, native language, culture, field dependence. For example, Gallagher, Bridgeman and Cahalan (2002) studied if the shift in testing format from traditional tests to computer-based assessment placed a disadvantage on any ethnic group or gender. The result shows no significant difference of one group over another.

Whichever kind of test teachers choose, they should not put themselves as the center of assessment. Teachers usually worry about the practicality, validity and reliability of the test and ignore other factors, e.g. students' anxiety that can lead to students' failure to do well on the test. These unpredictable factors that come into play when students are taking a test are what Bachman (1990) discussed, quoted in McNamara (2001). Bachman (1990) argued that the problem of the test is the misconception underlying test performance and learners' competence which might not be shown through their performance. And, the fact that the emphasis is solely put on performance can cause problems.

2.5.6 Computer-based tests and computer listening tests

Coniam (1998) suggested numerous points to consider when developing a computer-based listening test. Firstly, it has been stated that current listening tests fail to meet the criteria of 'authenticity' and 'interactiveness'. The tests also lack what Bachman and Palmer (1996) emphasized as 'test usefulness'

According to Bachman and Palmer (1996), today's computer-based listening tests' quality is still uneven. For example, the authenticity of the texts, the interactiveness of test format and the impact, including the washback effect are varied. Most important of all, the researchers found they lack construct validity. This underscores the poor construct defining and selecting processes. Also, it implies that computer-based test developers must give value to the process so that the test will contain the language abilities to be assessed.

A computer-based listening test, if created carefully, can be of great benefit. One of the most widely used computer-based proficiency tests is the TOEFL. Taylor et al (1999) studied the new response type included in the listening part of the computer-based TOEFL test and revealed that the difficulty of the listening comprehension item ranged in difficulty from .30-.96 with a mean difficulty of .66 using *p* value (proportion getting the item correct) as the index of difficulty. Regarding the reliability, the twenty items studied showed the estimated reliability at .78 as measured by coefficient alpha. The changes recorded are as follows. Firstly, testees have to click letters and pictures, sometimes presented in a diagram or a chart to answer questions as compared to the paper and pencil counterpart in which the format is MC. Secondly, they have to match or order information presented in a lecture or academic discussion. Moreover, they can control the volume, and their own progress through the test items. This increases the degree of control by test takers, which can lead to less anxiety. In the former test, students are given limited time to answer each question after the spoken text is delivered. Last but not least, there are pictures, photographs or diagrams incorporated to set the scene and to elaborate the content.

Dunkel (1991) explored a prototype of an EFL and ESL listening proficiency test. The format is still limited to MC, i.e. listening to a computer-generated segment of speech and related probe questions and answering by pressing number key 1, 2, 3 or 4. Test takers still have to choose the best answers from choices given; however, the multiple choice options are varied ranging from two to four. The test includes some

aspects of computer adaptive tests in which test takers are branched up and down based on their proficiency. The test consists of three sections.

1. understanding and making appropriate responses to questions
2. recognizing synonymous statements
3. comprehending monologues and conversations

The study was conducted with 19 students to explore the problems with hardware and software. The result showed their positive attitudes towards the test. There was only one student who did not like it. The test takers said they preferred the prototype test more than a traditional audio tape listening test. The study also revealed the comparable opinions towards this test and the TOEFL. The students felt that it was easier and more individualized than the TOEFL. Moreover, the text spoken was slower than that that in the TOEFL.

Most importantly, the study disclosed what teachers and test developers have to consider when writing a computer-based listening test.

1. The situations and test items should be varied and great in number.
2. The spoken text should include repetitions, and the speed should be slowed down.
3. The screen text should be printed with a consistent style of font, and the font should be large in size.
4. The test should contain more graphics.

In conclusion, a number of computer-based language tests have gained more attention from teachers and administrators. Guidelines of how to develop computer-based tests are, then, needed. To create a valid and reliable computer-based test, certain factors, including the interface design principles, must be considered. In addition, with its special features, the computer can be very beneficial to language assessment if used wisely.

In summary, Chapter II has reviewed the theories and the research studies relevant to what is focused on in this study. They include the listening ability and listening comprehension processes, the overview of learning strategies, especially cognitive and metacognitive strategies, the approaches of listening comprehension assessment, the issues concerning the integration of tasks in language tests, and the concerns over the development and the implementation of computer-based tests.

The literature review, particularly the part concerning cognitive and metacognitive strategies, shows that these two strategies show promising roles in helping enhance one's language learning and increase one's language performance. According to the work conducted by researchers such as Abraham and Vann (1987), Anderson (1985), Aek Phakiti (2003), Yu (2003) and Liu (2004), there are positive relationships between the use of the strategies and learners' performance. Also, other studies, such as in Chamot, Küper and Impink-Hernandez (1988), Oxford (1990), Vann and Abraham (1990) and Kaylani (1996), show different use of the strategies by the proficient group and the non-proficient group. These studies help establish the hypotheses of the study presented in Chapter 3, which discusses the research methodology of this research study.