Investigating the listening construct underlying listening-to-summarize tasks

Anchana Rukthong
Lancaster University, UK

Abstract

A major concern in language assessment is that test results or scores may not be generalizable to real-world language use, in view of the fact that test situations may be inherently different from more authentic settings. If this were the case, tests would probably yield invalid results. In this regard, this pilot study aims to shed light on the construct underlying listening-to-summarize tasks and the extent to which this task type can capture the processes used in real-world communication. Four Thai ESL learners participated in the study. They were asked to perform four listening-to-summarize tasks: two requiring an oral summary of listening and two requiring a written summary. Immediately after each task's completion, stimulated recall was conducted. The results show that the participants engaged in eight processes in their attempts to complete the tasks. These processes can be categorized into three main categories of processing: linguistic, semantic and discourse processing. The paper concludes that listening-to-summarize tasks can tap into the processes which are utilised by proficient listeners and necessary for successful academic listening.

Keywords: *cognitive processes, integrated test tasks, task-based language assessment,*

1. Introduction

Integrated test tasks, which include both receptive and productive language skills in task performance, are said to mirror academic literacy activity and reflect the ability to use language in authentic situations (Brown, Iwashita, & McNamara, 2005; Lewkowicz, 1997; Plakans, 2008; Plakans & Gebril, 2012; Weigle & Parker, 2012; Weir, 1990). Describing the construct underlying integrated test tasks is, however, not straightforward. Since at least two language skills are involved in task performance, it remains ambiguous what abilities are truly assessed by this task type and what abilities contribute to either success or failure in performance. In fact, when the underlying construct is not clearly understood or well defined, it is difficult for test developers to support their claims about construct representation and relevance and the usefulness of their tests. In this regard, the present study investigates the construct or abilities assessed by this task type by looking into the cognitive processes and sources of knowledge employed to complete the tasks.

To justify the meaning and value of test scores, it is crucial to study construct validity. A fundamental feature of construct validity, as discussed by Messick (1995), is construct representation, which might not be achieved only through relevant content and operative processes, such as examining the correlation of test scores with other external measures. A set of construct indicators, including cognitive processes, strategies and knowledge (metacognitive or self-knowledge) that are applied in task performance, is needed to explain the construct underlying test tasks (Messick, 1995). Thus construct validation has to take into account the meaning of test scores, not only in relation to test items but also to test takers and the context of assessment (Messick, 1995). This is needed in order to provide evidence and a rationale to support the trustworthiness of score interpretation and use (Messick, 1995). By investigating construct validity in this manner, research can explain the degree to which

interpretations and inferences made on the basis of test scores are appropriate and plausible (Bachman & Palmer, 1996; Chapelle, 1999; Messick, 1995; Xi, 2008). In addition, it can point to evidence that supports or discounts inferences or arguments made on the basis of test results (Messick, 1995).

The present study, in particular, focuses on the listening construct underlying listening-to-summarize tasks that include academic lectures as input. Listening is required in a variety of communicative events in academic settings, e.g. lectures, group discussions, tutorials, seminars and meetings with a supervisor. If students are to participate successfully in academic communication, they must have the ability to process and respond to spoken language (Lynch, 2011). Despite its importance, listening remains the least understood of the four language skills because of its ephemeral nature which is not directly observable (Buck, 2001; Field, 2013; Lynch, 2011; Rost, 2002). Though previous research has attempted to identify the construct underlying integrated test tasks (e.g., Brown et al., 2005; Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Frost, Elder, & Wigglesworth, 2011; Gebril, 2010; Gebril & Plakans, 2013, 2014; Plakans & Gebril, 2012), only a few studies aimed to investigate the construct of test tasks integrating a listening source text (e.g. Brown et al., 2005; Cumming et al., 2004; Frost et al., 2011). As for the studies that focus on integrated listening, they rely mainly on linguistic analysis of task performance to identify the test construct. None of them appeared to investigate participants' mental processes using stimulated recall. As acknowledged by language test educators (e.g. Bachman & Palmer, 2010; Messick, 1995), an investigation of cognitive processes utilized by the test takers is necessarily important in the description of the construct underlying the task due to the fact that it reveals thinking processes and knowledge used to complete the test tasks. Further research along this line is thus warranted.

Informed by Messick's (1995) construct validity, the study conceptualizes the construct underlying the listening-to-summarise tasks as the mental processes that test-takers engage in while performing the tasks. Data on test takers' cognitive processes is gathered and described using a cognitive process framework for listening. Literature related to the cognitive process framework and contributing factors of effective L2 listening is reviewed in the next section. Information concerning the research instruments, participants, data collection procedures and analysis is included in the research methodology section, followed by a discussion of research findings and a conclusion.

2. Literature Review

2.1 Cognitive process framework for listening

The description of the cognitive processes involved in listening-to-summarize tasks in the present study is framed by Field's (2013) cognitive processing framework. This model has been adopted for the following reasons. Firstly, it is a listening-based framework that takes into account both individual traits and the interaction between a listener and listening task, which is crucial when describing the listening construct (Buck & Tatsuoka, 1998; Rost, 2002). Secondly, this model has been established and modified on the basis of the processes used by proficient listeners in various contexts, including an academic environment, which is the context that the present study aims to generalize. Lastly, the model notes the role of higher-level processes which are required in real-world academic listening and which the tasks employed in the present study aim to tap into.

In this cognitive processing framework (see Figure 1), Field (2013) explains that successful listening performance entails five main levels of processing: 1) input decoding, 2) lexical search, 3) parsing, 4) meaning construction, and 5) discourse construction. These five

levels, as presented in the shaded boxes, are subdivided into lower-level processes and higher-level processes.

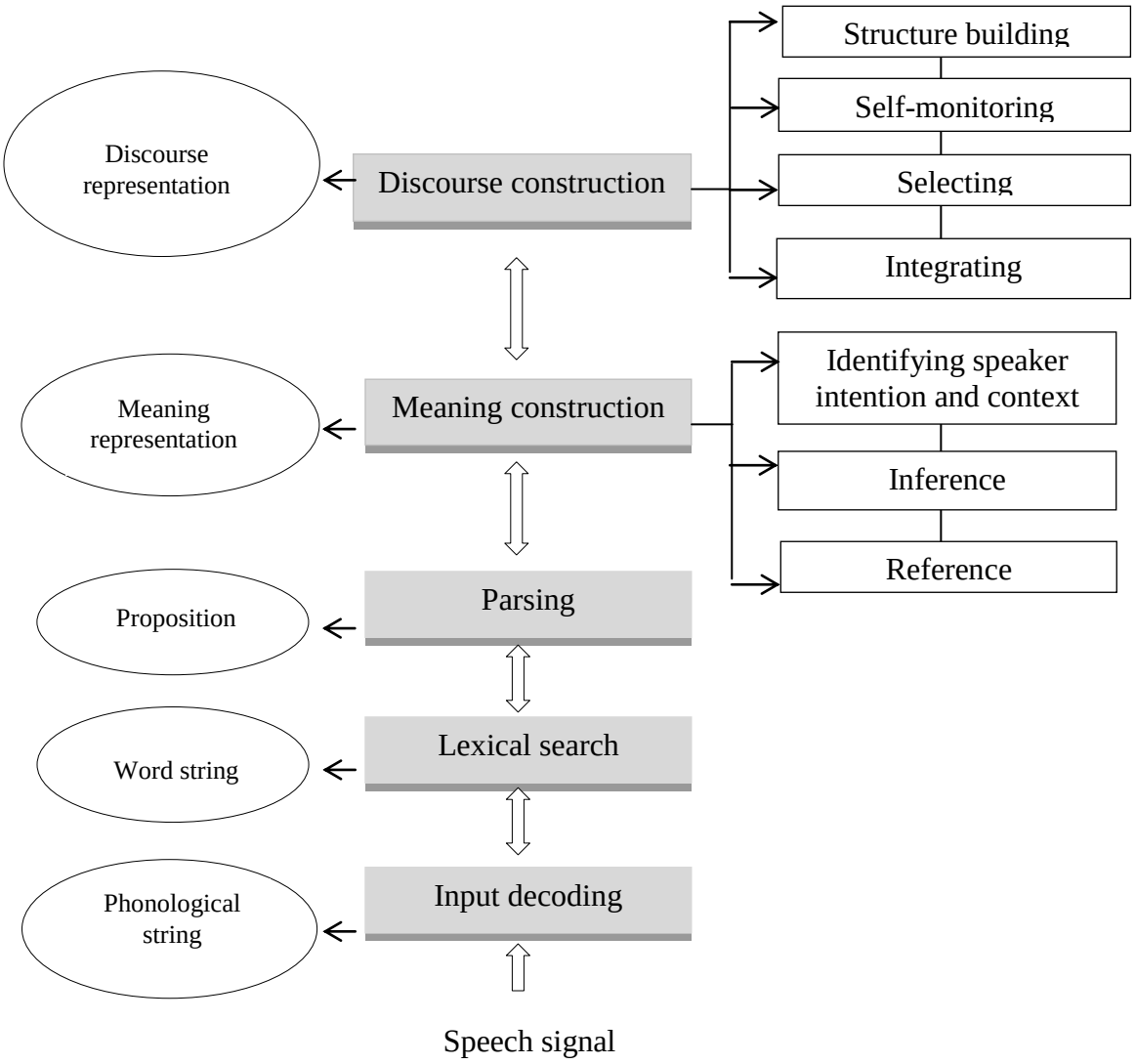

Figure 1: Field's cognitive processing framework for listening adapted from Field (2013)

Lower-level listening processes or linguistic processing involves the first three levels from the bottom (input decoding, word search and parsing), occurring when a message is being decoded into language. Higher-level processes, the top two processes, are associated with meaning and discourse construction. Although the processes are presented in a linear

order, it does not necessarily mean one stage of processing waits for one or more others. Language processes, as noted in this model, often act in a parallel and interactive manner. The numbering is thus used only to represent the levels. The oval shaped boxes in the figure indicate the output of each stage of processing.

Lower-level processes

Lower-level processes, according to Field (2013), involve three levels of linguistic processing, input decoding, word search, and parsing. Field (2013) indicates that listening processing starts from recognizing acoustic input and developing this to obtain a phonological string via input decoding, a set of words from lexical searching, and an abstract proposition via parsing. In input decoding, proficient listeners depend on their phonological knowledge to access a sequence of speech-like sounds and convert these sounds into representations that match the phonological system of the language being spoken (Field, 2013). At this level of processing, the listeners recognize a string of phonemes, some of which are marked as syllables of words. In a lexical search, the listeners map sounds to spoken word forms. Based on their lexical knowledge, the listeners have to determine word boundaries and identify words which are either content or function words in connected speech. At the level of parsing, the listeners segment units in the connected speech and construct propositions by applying their syntactic knowledge, understanding of standard word order, and intonation group boundaries.

Higher-level processes

Higher-level processes involve two levels of processing, meaning and discourse construction (Field, 2013). Listeners start to construct the meaning of what they have heard by relating the propositions obtained from lower-level processing, which is context-

independent, to their own schemata or concepts of knowledge they have developed. At this level, it is the task of the listeners to relate the propositions to the circumstances in which they were produced in order to extract their full meaning and relevance. The raw meaning of the speaker's words is often insufficient to convey the complete meaning of a text (Field, 2013). The listeners, therefore, have to supply additional information to comprehend what is said in a number of ways. One way to do this is to use pragmatic forms of language to interpret the speaker's intention. The listeners may also have to use contextual and semantic knowledge to relate propositions to the context in which they occur. The listener may, in addition, have to infer what the speaker left unsaid from what they have just heard or backtrack to what was being said or what was said earlier.

Discourse construction relates to four processes that the listeners apply to construct their understanding of a spoken text. As presented by Kintsch and van Dijk (1978), these processes are selecting, integrating, self-monitoring and structure building. Selecting is when the listeners assesse the relevance of an incoming piece of information, e.g. whether it is the repetition of a point made earlier or the central point of the topic being developed. On the basis of this consideration, the listeners may store the information being processed or discard it as irrelevant. Integrating is when the listener adds one or more new pieces of information to the discourse representation being developed. It involves recognizing conceptual links between incoming information and that already processed. Self-monitoring entails comparing whether a new piece of information is consistent with what has been processed before. If not, the listener has to consider whether the new judgement is correct or question whether what they have understood earlier and recall is correct. Structure building is when the listeners have to prioritize and organize the information they have stored according to its importance and relevance. A more proficient listener is able to build a more complex information structure than a less proficient one.

Field (2013) has pointed out that the processes described in this framework are built on the L1 listening comprehension processes. Successful listening, which means the listeners have a clear concept of what the speaker intends to say, depends not only on linguistic processing (input decoding, lexical search and syntactic parsing), but also on higher-level processes (meaning and discourse construction). While lower-level processes enable the listeners to produce propositions and understand the literal meaning of the message being conveyed, higher-level processes assist the listeners in relating the incoming message to their existing knowledge and building a knowledge structure, resulting in a complete understanding of the message. To achieve complete comprehension, the listeners must engage in both levels of processing. Higher-level comprehension is not really possible if lower-order processes are not working efficiently.

2.2 Factors contributing to effective L2 listening

L2 listening comprehension appears to be restricted by two main factors: the level of listener's knowledge and the level of expertise or automaticity in processing (Buck, 2001; Field, 2013; Rost, 2002). The knowledge involved in language processing concerns both linguistic and non-linguistic knowledge. Linguistic knowledge or language-related knowledge is a domain of information in the individual's memory, and it is available for use in tandem with metacognitive strategies to create and interpret discourse in language use (Bachman & Palmer, 1996). Language knowledge comprises two broad categories: organizational and pragmatic. Organizational knowledge includes grammatical knowledge, knowledge of vocabulary, knowledge of syntax and knowledge of phonology/graphology (Bachman & Palmer, 1996; 2010). In listening processing, this type of knowledge is employed mainly in linguistic processing. It enables the listener to encode speech into linguistic units, detect phonetic features and recognize words in connected speech in order to

interpret the incoming text. Pragmatic knowledge is generally activated at a high level of processing, i.e. meaning and discourse construction (Field, 2013; Kintsch & van Dijk, 1978; Rost, 2002). It entails functional and sociolinguistic knowledge, both of which enable the listener to interpret text discourse by relating utterances or sentences to the speaker's intention and to the characteristics of the language-use setting (Bachman & Palmer, 1996). Another type of knowledge that affects L2 listening is the cultural or world knowledge that the listener brings to a listening situation (Field, 2013; Rost, 2002). Such knowledge is shaped by the listener's cultural background and experience. Similar to pragmatic knowledge, cultural knowledge is activated mainly in high-level processing (Field, 2013; Rost, 2002). The listener has to apply this type of knowledge, especially when he or she has to make inferences or references relevant to the message being delivered in order to understand its full and essential meaning.

Effective L2 listening depends not only on the listener's knowledge but also the degree to which he or she can process knowledge automatically (Field, 2013; Rost, 2002). As indicated in the previous section, listening ability integrates a number of psycholinguistic abilities working in a parallel and interactive manner. Rost (2002) divides these into four levels: neurological processing, linguistic processing, semantic processing and pragmatic processing, which Field (2013) categorizes into two levels of processing based on the level of cognitive development, i.e., lower-level and high-level processes. Lower-level processes entail linguistic processing consisting of decoding, word search and syntactic parsing while higher-level processes comprise meaning and discourse construction. Under normal circumstances, linguistic or lower-level processes are considered to be fundamental to listening and the skills that must be acquired prior to the development of higher processes, i.e. semantic and pragmatic or discourse processes (Field, 2013).

L2 processing occurs in association with automatic and controlled processing (Shiffrin & Schneider, 1977). Automatic processes are cognitive processes that are well developed and which put little or no demand on processing capacity (Shiffrin & Schneider, 1977). They do not require conscious attention and are therefore unavailable to conscious awareness (Shiffrin & Schneider, 1977). Controlled processes, on the other hand, are conscious. They require attention and are used flexibly in changing circumstances (Shiffrin & Schneider, 1977). In fact, what is necessary for complete text comprehension is automatic word recognition and syntactic analysis (Goh, 2002). When these lower level processes become automatic, more cognitive capacity is freed up for higher-level processing, such as making references and inferences and constructing meaning (Goh, 2002). If there is difficulty in processing a message at the level of linguistic processing, such as sound perception or word recognition, language users have little cognitive capacity remaining for higher-level processing, resulting in incomplete comprehension (Goh, 2002).

In conclusion, the literature indicates that there is a major problem in the use of integrated test tasks to assess L2 performance, which relates to a lack of clarity concerning the abilities assessed by the tasks. Particularly in the case of tasks that include listening input, i.e. listening-to-summarize tasks, it remains unclear what listening abilities are performed and measured. One way to investigate the construct or abilities underlying the language test task, as acknowledged by language testers (e.g., Bachman & Palmer, 1996; 2010; Messick, 1995) is to look into cognitive processes test takers used during the task performance. Cognitive processing, as pointed out by listening researchers (e.g., Buck, 2001; Field, 2013; Rost, 2002), depends upon several types of knowledge, including both linguistic (e.g., phonological, lexical, syntactic, semantic, pragmatic knowledge) and non-linguistic knowledge (e.g., topical and world knowledge). In addition, cognitive processing has been found to vary from one listener to another, depending on their competence and abilities to use

such knowledge to process for understanding. Investigation of cognitive processes and sources of knowledge used during integrated task performance, thus, can point to the construct or abilities assessed this task type.

The present study attempts to describe the listening construct underlying integrated listening tasks, e.g. listening-to-summarise tasks, by answering the following research questions.

a) What cognitive listening processes do ESL test-takers engage in while performing academic listening-to-summarize tasks?
b) Are there any differences in the listening processes involved when different language modalities, namely speaking and writing, are required for summary production?

3. Research methodology

3.1 Listening-to-summarize task materials

The task materials used in the present study comprise four test items adapted from PTE Academic.[1] Two tasks require participants to orally summarize a listening passage (listening-speaking tasks) and the other two require a written summary of the listening text (listening-writing tasks). The listening-speaking tasks were adapted from the *Re-tell Lecture* items, which originally ask test takers to retell what they have heard. The listening-writing tasks were taken from the *Summarize Spoken Text* items. An image related to the content of the listening input was added to each of the two tasks. The purpose of the modification is firstly to make the four tasks investigated comparable in terms of what test takers are supposed to do, and secondly, to study whether different language modalities (speaking and

[1] Pearson Test of English Academic – an English proficiency test for non-native English speakers who need to demonstrate their academic English ability for university admission or professional purposes.

writing) have an impact on listening processes. The listening input for each task is 60 to 80 seconds.

Strictly following PTE Academic guidelines, the participants are allowed to listen once only but they can take notes while listening. For the listening-speaking tasks, after listening the participants have 10 seconds to prepare and then 40 seconds to give their oral summary. For each of the listening-writing tasks, participants have 10 minutes to write a 50–70 word summary of what they hear.

3.2 Participants

Four Thai students at Lancaster University (one undergraduate and three postgraduates) participated in a pilot study. For reasons of anonymity, they are referred to in the finding section as P1, P2, P3 and P4. Based on their performance scores, participants were categorized into two groups: moderate scoring participants (P1, P3 and P4) and a low scoring participant (P2).

3.3 Data collection procedures

Data were collected on a one-to-one basis in the following order:

1) Completion of a background questionnaire
2) Completion of two sample listening-to-summarize items (one listening-speaking task and one listening-writing task), in order to familiarize participants with the item type and reduce test anxiety
3) Completion of four listening-to-summarize tasks, i.e. two tasks requiring an oral summary (listening-speaking) and two tasks requiring a written summary (listening-writing). These tasks were presented to the participants as a PowerPoint presentation (PPT), which was timed and set to play automatically when

participants clicked on the start button. A counterbalanced design was employed in the task delivery with the aim of minimizing the effects of task sequencing on performance. With this design, the participants began the test with different tasks.

4) Participation in a stimulated recall immediately after each task completion. This was carried out in the participants' first language, in this case Thai.[2] In each recall which took about 15-20 minutes to complete, the participants were first presented with the video recorded during their task performance. Then they were invited to explain what they were paying attention to or thinking about while listening.

3.4 Data analysis

The data were analysed as follows.

1) ***Analysis of task performance.*** Two experienced human raters scored all the task responses, using the human rater version of the PTE Academic scoring criteria, in order to evaluate performance level. The oral summaries were scored on three aspects: content, pronunciation and fluency. The written summaries were marked for content, grammar, vocabulary, form and spelling.

2) **Analysis of stimulated recall data**. The stimulated recall data were transcribed and analysed to identify cognitive listening processes. Following the notion of qualitative data analysis suggested by Gass and Mackey (2000), the data were categorized into episodes. The data were first segmented into what appear to be plausible units that correspond to Field's (2013) cognitive listening processes (see Section 2.1). For example, the following extract, obtained from one pilot study participant's protocol, was segmented into two chunks.

> [Chunk_1] When I heard 'handicraft', I told myself that it was about hand-made stuff, // [Chunk_2] but then it [the audio-recording] didn't say anything about

[2]The quotes in the findings section are translated from Thai.

> items or products. Until I heard, 'his father', 'he', 'him', and 'the great scientist', I realized immediately that the listening was about a person [Hans Krebs], not about 'handicraft', as I had previously misunderstood.//

The first chunk (Chunk_1) indicates that the participant was trying to identify the word and its meaning she heard, which in this case she thought was 'handicraft'. This chunk was analysed to correspond to and categorized as lexical processing. The second chunk (Chunk_2) shows that the participants was trying to create a semantic relation between the words/phrases she heard in order to understand the real meaning of what she has been listening to. The participant linked 'his father', 'he', 'the great scientist' together and then realized that the audio was giving information about a person whose name was Hans Krebs, not about the hand-made as she understood in the beginning of the listening. This chunk corresponds to and was, thus, classified as semantic processing, i.e. reference making, in Field's framework.

4. Findings

Table 1 summarizes the cognitive listening processes demonstrated during the listening-to-summarize task performances obtained from stimulated recall data. Overall, the results show that the participants engaged in both lower- and higher- level processes to complete the tasks. Eight cognitive processes, in particular, were identified and categorized into three main types of processing: linguistic, semantic and discourse. Different language modalities (speaking and writing) required after the listening appeared to slightly affect the way the participants approached their listening tasks. Although the participants appear to employ the same processes, their performance scores show that they achieved different levels of success in their processing. Successful processing was found to depend, to a large extent, on the participant's linguistic and topical knowledge.

Table 1: Cognitive listening processes

Cognitive processes adopted by a proficient listener, as proposed by Field (2013)			**Processes demonstrated by the participants**			
Levels of processing	**Types of processing**	**Listening processes**	**Low scorer***	**Moderate scorers****		
			P2	**P1**	**P3**	**P4**
Lower-level processes	Linguistic	1) Input decoding	✓	✓	✓	✓
		2) Word search	✓	✓	✓	✓
		3) Syntactic parsing	✓	✓	✓	✓
Higher-level processes	Semantic	1) Identifying a speaker's purpose/context		✓	✓	✓
		2) Inferencing	✓	✓	✓	✓
		3) Referencing			✓	
	Discourse	1) Selecting				
		2) Integrating/linking pieces of information				
		3) Self-monitoring		✓		
		4) Structural building		✓		

*Average task performance score is lower than 40%

**Average task performance score is between 40% and 65%

1) Linguistic/lower-level processes

At the level of linguistic processing, three processes are found, namely input decoding, word search and syntactic parsing.

1.1) Input decoding

Input decoding, as described by Field (2013), is the lowest level of processing that takes place prior to word recognition. Words are recognized through the interaction of perceived sounds and context; and when listening to familiar words, this processing occurs automatically (Field, 2013). In this pilot study, only one participant (P1) explicitly indicated that he conducted input decoding. He describes, "here [the participant points to the video recorded while he was performing the task] I didn't know what the word was. I guessed from the sound I heard." The other three participants did not appear to engage in explicit input decoding. However, one can infer that they did use decoding, as they appeared to be able to

recognize words in connected speech as indicated in the next type of linguistic processing, i.e. word search.

1.2) Word search

All participants began their stimulated recall by describing that they were searching for key words while listening. For instance, P2 indicated, "I heard 'cells', 'human body', 'science' and 'study' and noted all the words down." P3 included in her notes, 'talent', 'really mean', 'high management', 'high ability', 'passenger happy' and 'I use the term to mean'. Then, from the words they recognized, they identified the points/ideas (propositions) in the listening task.

Although all the participants indicated that they started with word search when the listening began, it was found that moderate scoring participants (P1, P3 and P4) recognized words faster and more accurately than the low scorer, and as a result the moderate scoring participants were able to identify and infer the main points after listening to only a few sentences. P1, for example, said:

> When the listening started, I basically listened for vocabulary. Fortunately I recognized almost every word in the listening. I immediately understood it [the listening].

1.3) Syntactic parsing

The participants were found to adopt syntactic parsing for two purposes: to predict what was coming next in the listening and to build up propositions. Three participants (P1, P2, and P4) mention that they used syntactic parsing to predict what they were going to hear later.

P4 remarked:

> Here I was predicting that the speaker was going to talk about the definitions of talent because he said before, 'different ways of defining things restrictive, broad and meaningless'.

P2 reported:

> I know that he [the speaker] was going to talk about something in contrast because he said 'however… '.

In addition to syntactic parsing, all participants were found to make use of non-verbal information, (an image) provided as a supplement to the audio text, in order to recognize the words in connected speech and also to predict what the speaker was going to say next. For example, P1 commented:

> I predicted that the speaker was going to compare between the corruption and the income rate from the two graphs I saw while listening.

Two participants (P1 and P2) used syntactic parsing to build up their propositions. P1 explained:

> I have written in my note, 'as a result, ____ caused a disease'. I know that the missing word was a noun, so I put 'not having enough calcium in your blood' in the blank. It was a gerund phrase which I thought could function as a noun.

P2 used the word 'famous' in her proposition of 'he [Hans Krebs] is _____ for Krebs's cycle', because, as this participant said, "I know it needs an adjective and the adjective that often goes with for is 'famous'."

2) Semantic processes

Semantic processing, according to Field (2013), occurs when the listener is trying to understand the text beyond the literal meaning of the words uttered. It involves identifying the speaker's intention, inferencing and referencing, all of which were used by the participants.

2.1) Identifying speaker's intention

Field (2013) indicates that to comprehend a text, the listener also has to infer what the speaker leaves unsaid for whatever reason, e.g. believing it does not need to be included. The speaker's words are often insufficient to convey the full meaning of a message; and so the listener must identify what the real meaning of the message is. In this study, one participant (P1) was found to attempt to identify the speaker's intention in conveying a

message. P1 mentioned, "I think the speaker just wanted to add to the main point when he said 'no matter what parents say kids just do'."

2.2) Inferencing

On the basis of a few points that they could figure out, the participants started to identify the main points of the listening text. They did this by inferring from the words they noted down along with their background knowledge. For example, P2 stated:

> While listening, I had no idea what it [the audio text] was about at the beginning. Then, when I heard about 'cells', 'human body', 'science' and 'study', I assumed it was about a scientist.

However, it was found that the participant's topical knowledge caused one moderate scoring participant to misunderstand the story. As P4 stated:

> When I heard 'Hans Krebs' and 'he is a great scientist', I know that it was about a person. I studied about his life and work when I was in high school and I still remember his theory. I predicted that the story was about his life when he was a child and how he became famous.

It should be noted that, in the listening passage, Hans Krebs is only mentioned as an example of people who overcame obstacles and were successful in their life. This example is not the main point of the passage, but the participant misinterpreted it as the main point and thus scored "0" for content. This participant scored far less on this item than on the items she reported not having any content background on.

2.3) Referencing

Only one participant clearly demonstrated making use of referencing or linking reference words (e.g. he, him, this, what I just said, these factors) to their antecedents to construct the meaning of what was being said. This was stated by P3, one of the moderate scoring participants:

> When I heard 'handicraft', I told myself that it was about hand-made stuff, but then it [the audio] didn't say anything about items or products until I heard, 'his father', 'he', 'him' and 'the great scientist', I realized immediately that the audio text was about a person [Hans Krebs], not about 'handicraft', as I had previously misunderstood.

3) Discourse processes

Discourse processing occurs when the listener is constructing a discourse representation of what is being said. It is the highest level of listening comprehension processing (Field, 2013). Among the four types of discourse processing indicated by Field (2013), namely selecting, integrating, self-monitoring and structure building, the data obtained indicate only two types of processing, structure building and self-monitoring. Both of which were used only by P1, the moderate scoring participant who scored the highest in this study, in the two tasks that require an oral summary.

3.1) Structure building

The data indicated that P1 appeared to use this process only in the task that required an oral summary of listening. Because of the time constraints imposed by the oral summary task, this participant mentally outlined his summary while listening. He said, when watching his video, "here towards the end of the listening, I planned what I was going to say in the summary … from what I remember".

3.2) Self-monitoring

What is different between the highest-scoring participant (P1) and the lowest-scoring one (P2) is that, after they had predicted what the listening was going to be about, the lowest scorer (P2) listened and searched only for words she thought might help in constructing her predicted story. She was not aware that her predicted story could be wrong. The higher scorer (P1), though, was trying to predict and construct a mental outline, whilst also self-monitoring his own understanding by paying full attention to the rest of the listening, picking up on other key points and, realizing that his mental outline was not accurate, adjusting it. This participant (P1) said:

> At the beginning, I thought the speaker was going to describe the work of Hans Krebs so I planned to listen how Krebs's Cycle works, but then when the speaker

> mentioned 'obstacles' and 'example of how people have overcome difficulty in life', I realized that the point he [the speaker] was making was about how people became successful in life rather than the work of a famous scientist. As you can see (this participant pointed to his video), I went back and corrected what I'd noted down before.

In sum, the stimulated recall data reveal that the participants activated several types of cognitive processing while doing the listening-to-summarize tasks. The processes identified correspond to three main types of processing in Field's framework, i.e. linguistic, semantic and discourse. At the level of linguistic processing, listeners decode, word search and syntactically parse. Semantic processing involves identifying the speaker's intention, inferencing and referencing. Discourse processing includes self-monitoring and structure building. Two types of discourse processing, i.e. selecting and integrating, were not shown in the stimulated recall data. Although the findings are presented in sequential order, it should be noted that, during actual processing, most of the time these processes occurred in parallel and were interactive.

5. Discussion

The results suggest that it is possible for listening-to-summarize tasks to tap into the higher levels of cognitive processing while listening and the processes necessary in academic listening contexts. That is, as they appeared to perform successfully on the tasks, the participants in this study had to engage in meaning construction and discourse processing. According to Field (2013), these are higher-level processes used by proficient listeners and necessary for success in academic studies. Although the tasks allowed the participants to employ higher-level cognitive listening processes, the low scoring participant appeared to process mainly at the linguistic processing level, perhaps due to linguistic knowledge limitations.

The different modalities (speaking and writing) required after the listening tasks appeared to affect the listening processes slightly. That is, the oral summary component appeared to force one participant to engage in structure building while listening, whereas writing summary did not. This is evident in P1's recall statement. P1 reported that he was aware of the lack of time to construct what he was going to say before having to speak, so he was structuring the content of his summary towards the end of his listening task. In this case, it might be possible that if the participant had more time to prepare after the listening as it was in the case of the writing summary, he might not have involved in structural building. However, because of the unique characteristics of an oral summary task that allow less than one minute to prepare before speaking, the participant was forced to engage in structural building while listening. The other processes tapped into by the tasks are however, in general, quite similar.

6. Conclusion

This study has sought to describe the construct or abilities underlying integrated test tasks or, more specifically, listening-to-summarize tasks. Knowing what exactly this task type assesses is crucial for interpretations and inferences made on the basis of test scores obtained from this item type. By conceptualizing the construct underlying the tasks as cognitive processes used by test takers, the study has investigated and revealed the processes that test-takers engage in during task performance. The analysis of stimulated recall data shows that listening-to-summarize tasks tap into three main types of processing, corresponding to Field's framework, i.e. linguistic, semantic and discourse processing. Moderate scoring participants, categorized according to their overall performance scores, were found to engage more in higher-level processes, i.e. meaning construction and discourse processing. These processes, according to Field (2013), are used by proficient listeners and are necessary for success in

academic listening. The low scoring participant, however, was found to process mainly at the linguistic processing level, perhaps due to linguistic knowledge limitations. In further investigation, it would thus be useful to find out what processes determine individuals' success in task performance and what sources of knowledge, e.g. linguistic or non-linguistic knowledge, individuals mainly rely on to complete tasks successfully.

The findings presented in this study rely exclusively upon stimulated recalls conducted with four ESL Thai participants studying towards their postgrad studies. There appear some concerns regarding the use of stimulated recalls. That is, in some cases listening processes occur automatically and participants may not be aware of the processes they have used. Another concern is that by the time the participants had completed the summary, they might not be able to clearly think back to the way in which they processed the listening text. Regarding these, other research methods are highly recommended to supplement stimulated recall data in future research. One of the methods that could be useful is an analysis of the summary content produced by test takers as it has been acknowledged to reveal processes activated in task completion (Johns & Mayes, 1990). In addition, since this study involved a small number of participants in one particular context and only one coder was used in data coding, the generalization of the results to different L1 background and learning contexts should, thus, be done with care.

References

Bachman, F. L., & Palmer, S. A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks (TOEFL Monograph Series MS-29)*. Princeton, NJ: Educational Testing Service.

Buck, G. (2001). *Assessing listening*: Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157.

Chapelle, C. A. (1999). Validity in langauge assessment. *Annual Review of Applied Linguistics, 19*, 254-272.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 107-145.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening* (pp. 77-151). Cambridge: Cambridge University Press.

Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 1-25.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, N.J.: L. Erlbaum Associates.

Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing, 15*(2), 100-117.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly, 10*(1), 9-27.

Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing, 21*(0), 56-73. doi: http://dx.doi.org/10.1016/j.asw.2014.03.002

Goh, C. C. M. (2002). Exploring listening comprehension tactics and their interaction patterns. *System, 30*, 185-206.

Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of University ESL students. *Applied Linguistics, 11*, 235-271.

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

Lewkowicz, J. (1997). The integrated testing of a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7: Language testing and assessment, pp. 121-130). The Netherlands: Kluwer Academic Publishers.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes, 10*(2), 79-88.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13*(2), 111-129.

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing, 17*(1), 18-34.

Rost, M. (2002). *Teaching and researching listening*. Great Britain: Pearson Education Limited.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127-190.

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing, 21*(2), 118-133.

Weir, C. J. (1990). *Communicative language testing*. UK: Prentice Hall International Ltd.

Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education:* (2 ed., Vol. 7, pp. 177-196): Springer Science and Business Media LLC.