Investigating the Listening Construct Underlying Listening-to-Summarize Tasks

Anchana Rukthong

This thesis is submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics and English Language

Lancaster University

October 2015

Abstract

Integrated-test tasks, which combine receptive and productive language skills in task performance, e.g., listening-speaking or listening-reading-speaking, are increasingly being used in second language assessment, including in high-stakes English exams such as the TOEFL iBT and PTE Academic. Although recent studies (Plakans, 2008; Sawaki, Quinlan, & Lee, 2013) have found that the construct of each individual skill involved in task performance (e.g., listening, reading, and writing) is present and distinct, it is not entirely clear what abilities are actually assessed by the tasks, especially as far as listening is concerned. This study thus analysed test-takers' listening comprehension processing behaviours while completing listening-to-summarise tasks. In addition, test-takers' perceptions of the tasks and of listening task difficulty were investigated. The aim of this was to be able to describe the listening construct measured by integrated-listening tasks.

Data was collected from 72 Thai English as a Second Language (ESL) learners. Each participant completed four listening-to-summarize tasks – two tasks requiring an oral summary and two a written summary. To investigate the comprehension processing behaviours performed to complete the tasks, a stimulated recall was conducted with 12 participants after each task. To study the perceptions of the tasks and of listening difficulty and their relation to task performance, the remaining 60 participants completed a perception questionnaire after each task.

The results showed that to comprehend listening input with the aim of summarizing it, the participants engaged in both lower-level and higher-level cognitive processes and these cognitive processes were facilitated and monitored by a number of strategies. However, to maintain focus on the text's main point and accurately understand it, it was necessary that the participants successfully activated comprehension monitoring, real-time assessment of input, and lower-lever cognitive processes. Lack of the successful application of these processes and strategies often led to misinterpretations of the text, partly because of the interference of background knowledge which was not congruent with the texts' information. Participants with different performance levels were found to engage in different types of processes and strategies, with different degrees of success. The participants, in addition, were found to perceive the tasks as authentic and a fair way to assess their abilities to use English for academic purpose, especially listening abilities. In addition to providing a description of the listening construct measured by integrated-listening tasks, the study suggests that listening comprehension ability should be integrated in the description of the task construct and both cognitive and strategic processing should be recognized as part of the construct. On the basis of the findings, a model of second language (L2) listening in the context of listening-to-summarize tasks is formulated.

Acknowledgements

This thesis would not have been completed without the help and support of various individuals and organizations to whom I would like to offer my sincerest thanks.

First of all, I would like to thank Dr Tineke Brunfaut, my supervisor, for her constructive suggestions, enthusiastic encouragement, and valuable critiques throughout the process of completing this research. I very much appreciate her patience and willingness to give her time so generously. I would also like to thank Dr Luke Harding and Associate Professor Lia Plakans, as well as Professor Judit Kormos, for their constructive and insightful comments.

Additionally, I would like to thank Assistant Professor Dr Chonlada Laohawiriyanon, who has been a massive inspiration to me in academic life. Thank you for all the advice and opportunities she has offered to me and importantly for her belief in me. Without her support and encouragement, I would not have come this far.

Special thanks to Thai students of the 2013-2014 academic year at Lancaster and other universities in the UK, who, without hesitation, volunteered to participate in this research. Without their contribution, this research would not have been completed.

I would also like to thank my family and friends (both Thai and international). The amount of love and support from them has given me a lot of strength. Thank you also to John Heywood, a wonderful proofreader, and to Kathy and Dave, who have made me feel at home in Lancaster.

Lastly, I would like to thank Prince of Songkla University for its 50% PhD funding and Pearson Education Limited for its permission to use PTE Academic test items and the cooperation throughout my study. Without their support, the completion of this study would have been very difficult.

Table of Contents

Abstract ii

Acknowledgements iv

List of Figures x

List of Tables xii

Chapter 1 Introduction 1

1.1 Introduction 1

1.2 Background to the study 2

1.3 Scope of the study 7

1.4 Significance of the study 10

1.5 Definitions of terms 11

1.6 Structure of the thesis 13

Chapter 2 Literature Review 16

2.1 Introduction 16

2.2 Test validity 17

2.2.1 Earlier conceptualizations of test validity 18

2.2.2 More recent conceptualizations of test validity 19

2.3 Integrated-skills assessments 27

2.3.1 Defining integrated-test tasks 28

2.3.2 Previous research on integrated-test tasks 30

2.4 Frameworks for language comprehension processing 39

2.4.1 Anderson's (1985) cognitive framework for language comprehension 40

2.4.2 Rost's (2011) listening processing 44

2.4.3 Vandergrift and Goh's (2012) cognitive model of L2 listening comprehension 49

2.4.4 Field's (2013) cognitive framework for listening 52

2.4.5 Comparison of Field's (2013) listening comprehension model to Rost's (2011) and Vandergrift and Goh's (2012) models 55

2.4.6 Role of cognitive and metacognitive strategies in L2 comprehension processing 58

2.4.7 The model of listening comprehension processing employed in this study 61

2.5 Test-takers' perceptions of tasks and of listening task difficulty 68

2.5.1 Use of test-takers' perceptions in language testing........ 69

2.5.2 The investigation of test-takers' perceptions 72

2.6 Summary........ 76

Chapter 3 Methodology........78

3.1 Introduction........ 78

3.2 Research questions........ 79

3.3 Participants........ 79

3.4 Data collection methods and instruments 84

3.4.1 A background questionnaire 84

3.4.2 Listening-to-summarize tasks........ 85

3.4.3 Verbal protocols........ 90

3.4.4 Perception questionnaires 99

3.5 Pilot studies........ 101

3.6 Data collection procedures in the main study........ 107

3.7 Data analyses 113

3.7.1 Analysis of stimulated-recall data........ 113

3.7.2 Analysis of participants' notes and summary content 116

3.7.3 Analysis of test performance 120

3.7.4 Analysis of the questionnaire data 121

3.8 Summary........ 122

Chapter 4 Cognitive processes........123

4.1 Introduction........ 123

4.2 Cognitive processes 124

4.2.1 Acoustic-phonetic processing........ 127

4.2.2 Word recognition 128

4.2.3 Parsing........ 130

4.2.4 Semantic processing at the local level 133
4.2.5 Semantic processing at the global level 138
4.2.6 Pragmatic processing 144
4.3 Summary 147
Chapter 5 Cognitive strategies 148
5.1 Introduction 148
5.2 Cognitive Strategies 148
5.2.1 Prediction 151
5.2.2 Fixation 155
5.2.3 Inferencing 159
5.2.4 Elaboration 164
5.2.5 Reconstruction 169
5.3 Overall picture of the cognitive strategies used at each level of cognitive processing 180
5.4 Summary 182
Chapter 6 Metacognitive strategies 184
6.1 Introduction 184
6.2 Metacognitive Strategies 185
6.2.1 Preparing for listening 186
6.2.2 Selective attention 190
6.2.3 Directed attention 200
6.2.4 Comprehension monitoring 204
6.2.5 Real-time assessment of input 212
6.2.6 Comprehension evaluation 215
6.3 Summary 217
Chapter 7 Listening comprehension processing behaviours compared between tasks with different modalities and performance levels 220
7.1 Introduction 220
7.2 Participants' performance scores 220

7.3 Comprehension processing behaviours compared between performance levels and tasks with different modalities 224

7.3.1 Cognitive processes 224

7.3.2 Cognitive strategies 227

7.3.3 Metacognitive strategies 233

7.4 Overall picture of comprehension processing behaviours compared between intermediate and highly-able listeners during different periods of the listening tasks 240

7.5 Summary 245

Chapter 8 Perceptions of tasks and of listening task difficulty 247

8.1 Introduction 247

8.2 The participants' performance scores 248

8.3 The analysis of the questionnaires 249

8.4 Perceptions of task authenticity 250

8.5 Perceptions of test fairness 251

8.6 Perceptions of listening task difficulty 253

8.6.1 Code complexity 253

8.6.2 Cognitive complexity 257

8.6.3 Communication Stress 262

8.7 Relationship between perceptions of listening difficulty and task performance .. 265

8.8 Summary 279

Chapter 9 Discussion 280

9.1 Introduction 280

9.2 The processes and strategies used in comprehending listening input in listening-to-summarize tasks 280

9.2.1 Overall processes and strategies used to comprehend listening input 281

9.2.2 Role of strategies in listening task performance 290

9.2.3 Effects of task modality (speaking and writing) on listening processing 293

9.3 Role of listening comprehension and the listening abilities assessed by listening-to-summarize tasks 294

9.3.1 Role of listening comprehension 294

9.3.2 Listening abilities assessed by listening-to-summarize tasks 299
9.4 Insights into the validity of listening-to-summarize tasks 301
9.5 Summary 310
Chapter 10 Conclusions 312
10.1 Introduction 312
10.2 Summary of the study 312
10.3 Contributions of the study 321
10.3.1 Theoretical contributions 321
10.3.2 Methodological contributions 325
10.4 Implications of this study 326
10.4.1 Implications for test developers 326
10.4.2 Implications for language teachers 329
10.5 Limitations and future research 331
List of references 333
Appendices 341
Appendix 1: Ethics documentation 341
Appendix 2: Background Questionnaire 343
Appendix 3: Listening-to-speak task perception questionnaire 344
Appendix 4: Listening-to-write task perception questionnaire 345
Appendix 5: Coding scheme 346
Appendix 6: Table for data coding 350
Appendix 7: Examples of data 351

List of Figures

Figure 2.1: Links in interpretative argument by Kane et al. (1999, p.9) 21

Figure 2.2: The components of integrated-test tasks 30

Figure 2.3: Components of listening comprehension processing behaviors 63

Figure 2.4: Summary of processes and strategies and their sub-components................. 67

Figure 3.1: Sample of listening-to-speak task slides.. 88

Figure 3.2: Sample of listening-to-write task slides 89

Figure 3.3: An overall picture of the main study's data collection procedures 109

Figure 3.4: Task delivery and data collection 111

Figure 4.1: The participants' cognitive processes in understanding the listening input materials.. 126

Figure 5.1: Cognitive strategies used by the participants to solve listening problems .. 150

Figure 6.1: Metacognitive strategies used by the participants to monitor listening comprehension processing 186

Figure 7.1: Cognitive strategies compared between tasks with different modalities and between performance levels (*Corruption*) 228

Figure 7.2: Cognitive strategies compared between tasks with different modalities and between performance levels (*Talent*).. 229

Figure 7.3: Cognitive strategies compared between tasks with different modalities and between performance levels (*Vitamin D*) 231

Figure 7.4: Cognitive strategies compared between tasks with different modalities and between performance levels (*Hans Krebs*) 232

Figure 7.5: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Corruption*) 235

Figure 7.6: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Talent*).. 236

Figure 7.7: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Vitamin D*).. 237

Figure 7.8: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Hans Krebs*) 239

Figure 7.9: Comprehension processing behaviours compared between the highly-able and intermediate listeners during different periods in the listening tasks....................... 242

Figure 9.1: Listening comprehension model emergingfrom listening-to-summarize tasks 297

Figure 9.2: Listening abilities assessed by the listening-to-summarize tasks................ 300

List of Tables

Table 2.1: Previous research comparing the construct underlying integrated-test tasks and 32

Table 2.2: Previous research comparing performance characterisitcs between task types and performance levels 34

Table 3.1: Data on participants' backgrounds 83

Table 3.2: Participants' English proficiency demonstrated by IELTS scores 84

Table 3.3: Listening-to-summarizetasks 87

Table 3.4: Techniques used to collect verbal protocol data in previous studies investigating cognitive processes and strategies 95

Table 3.5: The perception questionnaire construct 101

Table 3.6: Research questions and an overview of data collection methods 107

Table 3.7: Analysis of inter-coder reliability 120

Table 5.1: Overall picture of cognitive strategies used at each level of cognitive processing 180

Table 7.1: Participants' performance scores 222

Table 7.2: The participants whose listening behaviours were compared for each listening topic and between tasks with different modalities 223

Table 7.3: The cognitive processes compared between performance levels and between tasks with different modalities 225

Table 8.1: Group B participants' performance scores 248

Table 8.2: The analyses of questionnaire components and reliability 249

Table 8.3: Perceptions of task authenticity 251

Table 8.4: Perceptions of tasks fairness 252

Table 8.5: Perceptions of code complexity (*Corruption*) 255

Table 8.6: Perceptions of code complexity (*Hans Krebs*) 255

Table 8.7: Perceptions of code complexity (*Talent*) 256

Table 8.8: Perceptions of code complexity (*Vitamin D*) 256

Table 8.9: Perceptions of cognitive familiarity (*Corruption*) 258

Table 8.10: Perceptions of cognitive familiarity (*Hans Krebs*) 258

Table 8.11: Perceptions of cognitive familiarity (*Talent*) 258
Table 8.12: Perceptions of cognitive familiarity (*Vitamin D*) 259
Table 8.13: Perceptions of cognitive processing demand (*Corruption*) 260
Table 8.14: Perceptions of cognitive processing demand (*Hans Krebs*) 260
Table 8.15: Perceptions of cognitive processing demand (*Talent*) 261
Table 8.16: Perceptions of cognitive processing demand (*Vitamin D*) 261
Table 8.17: Perceptions of communication stress (*Corruption*) 264
Table 8.18: Perceptions of communication stress (*Hans Krebs*) 264
Table 8.19: Perceptions of communication stress (*Talent*) 264
Table 8.20: Perceptions of communication stress (*Vitamin D*) 265
Table 8.21: The relationship between the perceptions of task difficulty and task performance (*Corruption*: Listening-to-speak task) 268
Table 8.22: The relationship between the perceptions of task difficulty and task performance (*Corruption*: Listening-to-write tasks) 268
Table 8.23: The relationship between the perceptions of task difficulty and task performance (*Hans Krebs*: Listening-to-speak task) 270
Table 8.24: The relationship between the perceptions of task difficulty and task performance (*Hans Krebs*: Listening-to-write task) 271
Table 8.25: The relationship between the perceptions of task difficulty and task performance (*Talent*: Listening-to-speak) 273
Table 8.26: The relationship between the perceptions of task difficulty and task performance (*Talent*: Listening-to-write) 274
Table 8.27: The relationship between the perceptions of task difficulty and task performance (*Vitamin D*: Listening-to-speak) 276
Table 8.28: The relationship between the perceptions of task difficulty and task performance (*Vitamin D*: Listening-to-write task) 276

Chapter 1 Introduction

1.1 Introduction

Integrated-test tasks, which combine at least two language skills in task performance, such as reading-writing and listening-reading-speaking, have been included in assessments of second language proficiency for at least a decade. This seems a response to awareness that language communication in the real-world rarely involves the use of one language skill in isolation but is usually a combination of language skills, e.g., listening-speaking. However, since at least two language skills (modalities) are involved in integrated-test task performance, i.e. a receptive skill (listening/reading) and a productive skill (speaking/writing), it remains unclear what abilities are actually assessed by this task type. Research has been conducted to investigate its underlying construct. However, while a body of research aimed to describe the construct of the productive skills (speaking/writing) integrated in this task type, only a small number of studies have attempted to investigate comprehension ability as far as integrated-listening tasks are concerned. In fact, the credibility of language tests depends, to a great extent, on a clear description of the test construct or the abilities the test assesses. When the test construct is not clearly understood or well defined, it is difficult for testers to justify interpretations and decisions made on the basis of the test scores. Therefore, this study set out to uncover the construct underlying the receptive skill (listening), in integrated listening-to-summarize tasks.

In this chapter, first the background (1.2) and the scope (1.3) of the study are described. Next, its significance is introduced (1.4) and definitions of key terms are provided (1.5). The last section (1.6) outlines the thesis's structure.

1.2 Background to the study

A major concern in communicative language assessment is the extent to which tests used tap into the abilities to use language beyond the test situation or, more specifically, in real-world communication. Test results (scores) may not be generalizable to real-world language use since test situations may be inherently different from more authentic settings and tests may not capture abilities that represent the demands of language use in authentic communication (Norris, Bygate, & van den Branden, 2009). If this were the case, interpretations and inferences made on the basis of test scores would be invalid and tests would not be useful (Elder, Iwashita, & McNamara, 2002).

Since acts of real-life communication often require at least two or more language skills, such as listening-speaking, integrated-test tasks which require test-takers to perform language skills (listening, speaking, reading, and writing) not in isolation but in an integrated manner, such as listening-speaking or listening-reading-writing, have been introduced and are used widely in language tests (Cumming, 2014). Specifically, this task type has been adopted in high-stakes tests of English for Academic Purposes, such as the Test of English as a Foreign Language (TOEFL iBT: for sample items see http://www.ets.org/c/17722/audio/vol_3/track10vc180396.mp3) and the Pearson Test of English (PTE) Academic (https://www.youtube.com/watch?v=z6jPSFKCejQ&index=4&list=PLlwSacA9VMa7ubmtbLGeUnBmvIu-vKYoW).

The rationale underlying the use of integrated-test tasks in second language assessments, as can be concluded from previous research, relates to at least four benefits of this task type. First, as perceived by experienced participating teachers in Cumming, Grant, Mulcahy-Ernt, and Powers (2004), integrated-test tasks represent characteristics of real-world tasks. The tasks, in addition, and found in Asencion (2004) have the potential to capture abilities required in actual academic contexts, some of which may not be tapped into by independent-skill tasks.

The second benefit of integrated-test tasks relates to scoring reliability. Weigle (2004) found that when comparing two conditions of a writing test – one allowing test-takers to write based solely on their experiences and background knowledge and the other requiring test-takers to write a summary of a given text – the latter condition led to higher scoring reliability; both greater scoring consistency and a high level of agreement on score points between the raters were obtained. Weigle explained that this was because the source text suggested necessary or desirable content that should be included in the essay, on which the raters relied when scoring the task performance.

The third advantage of integrated-test tasks concerns the potential of the tasks to promote equality and fairness. Cumming et al. (2004) observed that in an independent-writing test task, where only a topic was given as input, students with knowledge of the topic were able to write fluently while those without it were not. The students' background knowledge, as the researchers stated, had to provide ideas for writing; thus any lack of such knowledge had implications for the writing. In contrast, in a reading-writing task, the students had to read the input material and generate ideas from the reading text. Although background or topical knowledge still played a role in reading comprehension, the researchers found that it did so less in the integrated-test tasks. This is because in the latter tasks, the students did not have to write by relying solely on their topical or background knowledge. Based on the findings, the researchers concluded that provision of the source text, to some extent, put students on an equal footing in terms of providing content and stimulating ideas to write about under test conditions.

Fourth, the use of integrated-test tasks has been observed to have positive washback or positive impact on language teaching. Weigle (2004) indicated that as a result of the introduction of integrated reading-writing test tasks in a university test of English for non-native speakers, classroom instruction changed from teaching writing in isolation to focusing more on practising writing in combination with reading such as writing based on reading materials. Particularly, students were shown how to critically analyse source materials and appropriately integrate the texts into their writing. These abilities, as Weigle pointed out, are necessary not only to achieve success in the tests but also on academic courses.

Despite these advantages, the use of integrated-test tasks in language assessment has encountered a great challenge when test validity is of concern. Integrated-test tasks have been criticised for their potential effect of task dependence, making it difficult to know what construct is being assessed by the tasks. Weir (1990) refers to this effect as ‘muddied measurement’ and explains that it occurs when performance on one item interferes with performance on a subsequent item. In the case of integrated-test tasks, this could happen, for example, when performance in speaking or writing depends on successful comprehension of a source text. When test-takers perform poorly, it is, thus, difficult to determine what ability or knowledge weaknesses may have caused this (Lewkowicz, 1997). For example, in the case of a reading-based summary-writing test, it is not easy to know whether a poor summary in which some of the main points are missing is due to poor comprehension or poor writing skills (Alderson, Clapham, & Wall, 1995).

Having clear test constructs or knowing what abilities tests actually measure is crucial in evaluating test validity or demonstrating the extent to which tests accurately assess test-takers' abilities, as well as judging whether the interpretations made on the basis of test scores are meaningful and valid. This is especially crucial in performance-based assessment, where tests are often assumed to assess the intended abilities by the use of communicative tasks thought to represent characteristics of real-world tasks (McNamara, 1996). Despite the use of these tasks, Bachman (2002) and Weir (2005) point out that, in practice, test performances are affected by several factors, including test-takers' personal factors (e.g., background knowledge and test anxiety), task factors (e.g., task difficulty), and task administration. Abilities assumed to be assessed by the test tasks may not actually be assessed. Thus, the use of 'so-called' authentic tasks (representing characteristics of real-world tasks) may, in reality, not be authentic if they do not assess abilities required in the target situation (Bachman, 2005; McNamara, 1996; Weir, 2005).

It is important that test constructs, as Weir (2005) contends, are defined at an initial stage of test design and development to guide the selection of appropriate test tasks. As the constructs are normally defined according to related theories, they are referred to as theoretical constructs (Weir, 2005). To ensure that tests assess the desired constructs/abilities, the theoretical constructs, have to be checked against actual test performance, for example, by exploring test-takers' cognitive processes, as Messick (1989) and Weir (2005) suggest. If so, the interpretations and decisions made on test scores can then be justified. In the case of integrated-listening tasks, it is not clear what construct this task type is supposed to measure although the tasks are considered useful in many ways, e.g., for representing characteristics of real-world tasks. This could make it difficult for testers to first decide whether the tasks would meet their testing purpose and then to interpret the test scores. As a consequence, an investigation of what this task type assesses is necessary, and on the basis of this, the theoretical construct of this task type can be formulated.

In conclusion, the use of integrated-test tasks in L2 assessments has been observed to have a number of benefits, including the potential for achieving greater task authenticity and tapping into abilities to use language in real-life communication. In addition, the inclusion of input materials in this task type has been shown to diminish issues related to test-takers' background knowledge and topic familiarity in test situations. The content of integrated-test responses has been found to be scored more consistently than that of independent-skill tasks. The tasks have also been found to initiate positive washback in classroom settings. However, describing the construct underlying integrated-test tasks is not straightforward. Due to the combination of at least two language skills in task performance, it remains ambiguous what abilities are truly being assessed by this task type and what abilities contribute to either success or failure in performance. In fact, the credibility (overall quality) of language tests depends, to a great extent, upon a clear description of the abilities or the construct they measure. When the underlying construct is not clearly understood or well defined, it is difficult for test developers to select tasks that suit their testing purpose and support their claims about construct-representation and -relevance and the usefulness of their tests. In this regard, an investigation of what abilities tasks measure seems crucial to inform the construct underlying integrated-test tasks.

1.3 Scope of the study

On the basis of the issues outlined in the previous section, this study aims to investigate the listening construct or abilities assessed by integrated-listening tasks. Specifically, the study will attempt to define the listening construct of listening-to-summarize tasks that include an academic lecture as input. Following Messick's (1989) unified concept of test validity and Weir's (2005) cognitive framework for test development and validation, the study conceptualizes a test construct as the cognitive processes and strategies test-takers engage in during task performance and investigates 1) test-takers' cognitive processing for listening comprehension and 2) their perceptions of tasks and listening task difficulty. This is in order to define the listening construct measured by the tasks, the context-appropriateness of integrated-test tasks (as perceived by test-takers).

Although listening-to-summarize tasks involve two skills in task performance (listening and either speaking or writing), listening will be the primary interest in this study for a number of reasons. First, it has been suggested in the literature related to integrated-skills assessments and language testing in general that listening is the least researched skill compared to other language skills (writing, speaking, and reading), hence it is the least known in terms of its construct (Buck, 2001; Field, 2013; Lynch, 2011; Rost, 2011). Second, studies on integrated-test tasks have mainly concentrated on the productive rather than receptive skills (see 2.3.2). Such studies often compared the characteristics of performances on independent-writing tasks to those on integrated reading-writing tasks to investigate whether these two task types assessed the same construct. Third, in previous research there is no consensus on the role of comprehension ability involved in task performances. Whereas studies such as Cumming et al. (2004) and Gebril (2010) concluded that independent-skill tasks and integrated-skills tasks measure the same construct, others such as Asención Delaney (2008) and Plakans (2008) indicated that these two task types measured different constructs and both the productive skills and receptive skills involved in task performances are assessed by the integrated task type. Fourth, although it is likely that comprehension plays a role in integrated-listening task performance, little research has paid attention to this area. To address these research gaps, this study aims to investigate the listening abilities assessed by integrated-skills tasks, namely listening-to-summarize tasks, in order to describe the role of listening comprehension (if any) and the listening construct underlying the tasks.

The tasks investigated in this study include an academic lecture as listening input. Academic listening is of particular interest for two major reasons. First, listening is required in a variety of communicative events in academic settings, e.g., lectures, group discussions, tutorials, seminars, and meetings with a supervisor. The ability to understand and respond to academic listening is required for students to participate successfully in academic communication and in academic success in general (Lynch, 2011). Second, lecture listening in particular, as pointed out by scholars such as Buck (2001), Lynch

(2011), and Taylor and Geranpayeh (2011), constitutes complex processing and so far it has been unclear what successful lecture comprehension entails. Listeners, as Buck (2001) stated, can fail to comprehend a text's main point despite their understanding all the words in the text. Taylor and Geranpayeh (2011) explain that this is because academic listening is generally cognitively demanding, especially in terms of the amount of information listeners have to process simultaneously. Academic texts additionally are generally context-reduced, requiring the listeners' logic and inferences to understand the points being delivered (Taylor & Geranpayeh, 2011). Despite these complexities, little research has attempted to reveal the construct underlying lecture listening. Further research focusing on this area is thus warranted.

In practice, listening-to-summarize tasks with different modalities, i.e. listening-to-speak (requiring an oral summary after the listening) and listening-to-write (requiring a written summary after the listening) will be investigated. These tasks are of interest because they are considered to tap into high-level processing abilities which are crucial for success in academic listening but may not be captured by other forms of response, e.g., multiple-choice and gap-filling questions (Field, 2012; Taylor & Geranpayeh, 2011). In addition, summaries (both in oral and written forms) required after the listening are believed to have potential for tapping into processes such as meaning building and discourse construction (Johns & Mayes, 1990; Kintsch & van Dijk, 1978). These are all required in academic studies and therefore important to be assessed by tests in order to fully represent the construct of academic listening.

1.4 Significance of the study

This study hopes to advance our understanding of the listening construct underlying integrated-listening test tasks, i.e. listening-to-summarize tasks. The research findings will describe the set of cognitive processes and strategies employed during listening task performance, including the sources of knowledge test-takers use to comprehend listening input. In addition, the study intends to reveal test-takers' perceptions of tasks and listening task difficulty and the extent to which the listening task difficulty, as perceived by test-takers, relates to their listening performance.

Practically, this study hopes to provide information that can assist test developers and item writers, when considering the use of listening-to-summarize tasks in their context. Specifically, it will give them an idea of what listening abilities these tasks assess and to what extent the tasks, as perceived by test-takers, can represent characteristics of real-world tasks. This is to justify their use of listening-to-summarize tasks.

In relation to second language (L2) teaching, the study hopes to extend the conceptual understanding of what academic listening involves and what abilities or knowledge students need in order to successfully understand lectures on various topics. Academic texts, as described earlier, are unique and require listeners' inferences and logical thinking to be understood and it is also not fully understood how individuals comprehend such texts (Buck, 2001; Taylor & Geranpayeh, 2011). By looking into listeners' task processing behaviours, this study will reveal how listeners approach their listening, what sources of knowledge they rely on for text comprehension, and what could make them either succeed or fail in their lecture listening. In addition, the study will point to strategies that help listeners achieve their listening goals and that might have to be focused on in L2 teaching.

Theoretically, the study will provide empirical evidence to inform the theoretical construct of cognitive processing for listening comprehension. One influential framework that has recently been introduced and used to analyse test-takers' cognitive processes for listening is Field's (2013). However, this framework was drawn up from data provided by first language (L1) and competent L2 listeners. Although Field's model is considerably comprehensive and describes clearly what is involved in successful listening processing, it may not describe those behaviours performed by listeners with lower ability who can also be successful. Unlike Field's (2013) model, this study attempts to describe the listening processing behaviours of advanced L2 listeners in comparison with those of intermediate L2 listeners. The investigation of test-takers' processes in this study, although in a simulated testing situation, is hoped to fine-tune the existing cognitive processing framework for listening in language assessment in general and for integrated-test tasks more specifically.

1.5 Definitions of terms

Throughout this thesis a number of terms will be used. This section presents definitions of terms that are key to this study to avoid confusion as some of these terms may be used differently in other studies. Here follow the terms.

'Skills', defined after Richards and Schmidt (2002), is 'the mode or manner in which language is used (p.293)'. It refers to four language skills, i.e. listening, speaking, reading, and writing.

'**Abilities**' are defined as individual capacities to perform an act (Davies, 2005). In this study the term is used in combination with language skills to refer to the cognitive processes and strategies that underline specific skills. For example, listening abilities involve abilities to decode sounds, segment information in continuous speech, make inferences, etc.

'Cognitive processes' are a category of mental actions that contribute directly to text comprehension (Anderson, 1985; Rubin, 1981). Following Shiffrin and Schneider (1977), cognitive processes are limited to processes that are well developed and put little or no demand on processing capacity.

'Strategies' are mental actions that are purposefully activated by users. Defined after Shiffrin and Schneider (1977), strategies are controlled processes, which require attention and are used flexibly in changing circumstances. The use of strategies in text comprehension processing is categorised into two groups: cognitive strategies used to solve comprehension problems; and metacognitive strategies used to manage cognitive processes and strategies.

'**Tasks**' is used to refer to work assigned to language learners/test-takers to complete for some purposes (Candlin, 2009). Tasks provide learners/test-takers with language input and require them to produce task output on the basis of their comprehension of input materials (Candlin, 2009). In language testing, tasks are used to elicit test-takers' abilities for the purpose of evaluation.

'**Task types**' is classified according to a number of language skills involved in task performance. Two types of tasks are referred to in this study: 1) independent-skill tasks, where only one language skill is involved, e.g., independent-speaking tasks, and 2) integrated tasks, where at least two language skills are involved, e.g., integrated listening-speaking tasks.

'**Tasks with different modalities**' refers to the integrated-skills tasks that provide the same input material but require different skills in task production. In this study, the term is used to refer to the listening-to-write and the listening-to-speak tasks.

1.6 Structure of the thesis

This thesis is composed of ten chapters. Following this introduction chapter is the literature review (Chapter 2). In this chapter, first different conceptualizations of test validity and methods for validation are reviewed in order to situate this study in a validation framework. This initial section argues that although several types of evidence are needed to support test validity, evidence of test constructs or the abilities tests measure should take priority because they form the basis for the interpretation and use of scores. Next, the characteristics of integrated-test tasks which are the focus of this study and previous studies on this task type are presented. At the end of this section, the justifications for this study to investigate the listening construct underlying integrated-listening tasks are provided. The remaining section of this chapter reviews the literature on comprehension processing frameworks and listening in particular to guide the analysis of test-takers' listening processing behaviours. Finally, the importance of test-takers' perceptions in test validation is considered. This is to frame the investigation of test-takers' perceptions of task characteristics that might affect listening performance.

Chapter 3 describes the research methodology. First, the research questions are formulated and the research participants are described. Next, the data collection methods and research instruments, comprising a background questionnaire, listening-to-summarize tasks, verbal protocols, and a perception questionnaire are explained. To try out the research instruments and to investigate the feasibility of the research design, three pilot studies were conducted and are described in this chapter. After that, the data collection and data analyses of the main study are presented.

The findings are presented in five consecutive chapters. The first three chapters (4-6) present the findings of the test-takers' comprehension processing, which inform Research Question 1 (RQ 1). Specifically, Chapter 4 presents the findings regarding cognitive processes; Chapter 5 cognitive strategies, and Chapter 6 metacognitive strategies. In these chapters, the overall picture of the processes/strategies the test-takers engaged in is first presented, followed by the description of each individual process/strategy. Chapter 7 compares the processes and strategies used by the test-takers between tasks with different modalities and performance levels (RQ 1a). Chapter 8 presents the findings of perceptions of tasks and listening task difficulty addressed in RQ 2 and 2a by first presenting the perceptions of task authenticity, fairness, and listening task difficulty. Then the relationship between participants' perceptions of listening task difficulty and task performance are provided.

In Chapter 9, all these findings are then discussed in association with the research questions and the literature. The study is concluded in Chapter 10 by first summarizing the study and its main findings. Then the contributions and implications of the study are outlined. In the final section, the limitations of the study and recommendations for future research are indicated.

Chapter 2 Literature Review

2.1 Introduction

With the aim of providing empirical data on the listening construct underlying listening-to-summarize tasks, a number of works related to the focus of the study are reviewed here. First, in section 2.2, conceptualizations of test validity and validation frameworks are reviewed to position this study in the field of language assessment. It is argued in this section that having a clear test construct is important in a test validation process and thus, in tests where the construct is not clearly defined, more research on the construct or abilities measured by the test tasks has to be carried out. Next, in order to provide an understanding of the tasks in question, the theoretical concepts of integrated-skills assessments, the advantages and limitations of integrated-tasks, as well as the previous research conducted on the use of this task type are reviewed in section 2.3. This is also to identify gaps in the previous research and provide the rationale behind this study.

Since the test construct in this study is conceptualized as the cognitive processes and strategies test-takers activate while performing test tasks, data on these cognitive processes and strategies will be gathered and analysed to define the listening construct. Consequently, section 2.4 reviews existing frameworks related to language processing in general and listening processing more specifically, in order to provide guidelines for the analyses of listeners' processing behaviours in this study. First, Anderson's (1985) framework is described to provide an understanding of what comprehension processing generally entails (2.4.1). Then, two influential frameworks for listening comprehension processing in the second language acquisition context, i.e. Rost's (2011) and Vandergrift and Goh's (2012) are reviewed (2.4.2-2.4.3). After that, Field's (2013) cognitive framework for listening processing, which has been introduced in the language testing context and which is relevant to the context of this study, is described (2.4.4). All these models are then compared and this leads to a description of the role of cognitive and metacognitive strategies in L2 comprehension processing (2.4.5). This section ends with the presentation of the framework used to analyse test-takers' listening processing behaviours in this study (2.4.6). Section 2.5 reviews the literature on the role of test-takers' perceptions in test validation and factors that might affect task performance to guide the investigation of test-takers' perceptions. The chapter concludes with summarising the gaps in the literature and suggesting the aspects to be addressed by this study (2.6).

2.2 Test validity

A central problem in language assessment is to what extent tests in general accurately assess test-takers' language abilities or more specifically, to what extent current tests used are valid. Test validity, according to testing scholars such as Alderson et al. (1995), Bachman (2005), and Weir (2005), has become a key issue for all language testers since it defines the quality of tests. According to these authors, tests would not be useful if they lacked validity. Interpretations or predictions of test-takers' abilities made on the basis of test scores, furthermore, are unlikely to be justified without evidence of test validity (Bachman, 2005). Therefore, to prove the usefulness of tests, validation, or the process of assessing test validity, has to be carried out. Over time, however, there have been different conceptualizations of test validity and these have been associated with different methods of test validation. The following sections thus describe the different conceptualizations of test validity and the ways in which test validity has been investigated. The ultimate purpose of this section (2.2) is to point out the significance of this study in relation to test validation.

2.2.1 Earlier conceptualizations of test validity

In earlier views on validity, such as those presented in Cronbach and Meehl (1955) and Lado (1961), test validity is conceptualized as consisting of several independent types of validity. Based on these views, different types of test validity were identified and researched as part of the validation process. Cronbach and Meehl (1955), for example, distinguished four categories: predictive validity, concurrent validity, content validity, and construct validity. The first two types, as Cronbach and Meehl explain, are criterion-oriented. When tests provide results that show a relationship with other measures (criteria taken on the same test-takers and given at the same time), they are considered as having concurrent validity. If tests show a relationship with results on other tests (measures) given at a later point in time, they are considered as having predictive validity. Alderson et al. (1995) later refer to these two types of validity as external validity. Content validity, on the other hand, refers to the representativeness of the content tests are meant to represent. Construct validity, according to Cronbach and Meehl (1955), concerns the extent to which tests actually measure the construct or theoretical definition of language abilities that is aimed to be measured. Additionally, in Alderson et al. (1995), two more types of validity are added: face validity and response validity. Face validity relates to the public acceptability of a test as a proper test (Alderson et al., 1995), whereas response validity concerns the extent to which test-takers' actual response processes reflect the expected ones.

These types of validity, according to these earlier views, are independent from each other. As a consequence, test validation has looked into different types of validity, depending on what qualities of the tests testers are interested in. Content validity, for example, is investigated when testers want to know whether a test's content represents the knowledge structure the test is meant to be concerned with. Or testers may want to study predictive validity in order to know whether tests can accurately predict future performances of test-takers. Construct validity is carried out to investigate whether a test construct is reflected in the scores obtained. Several methods have been used to investigate test validity in these views, including for example correlation analyses between test scores and scores obtained from other measures or analyses of test specifications in comparison to expert judgments. Although different methods are used, test validation in these earlier conceptualizations, as Xi (2008) contends, is to support score-based predictions rather than score-based interpretations.

2.2.2 More recent conceptualizations of test validity

Alternative views on validity have been formulated particularly since the introduction of Messick's (1989) unified concept of test validity. In Messick's view, test validity not only accounts for the quality of a test itself but is a characteristic of the inferences drawn on the basis of test scores and the consequences of the assessment as a whole (Messick, 1995). Test validity should not only be demonstrated through relevant content and statistical analyses such as correlations between test scores and other external measures.

Instead, validity concerns the extent to which tests can be shown to produce data, i.e. test scores, which are accurate representations of candidates' levels of language knowledge or skills in relevant contexts (Messick, 1995).

A primary concern of test validity, as postulated by Messick (1989), is construct representation. Although tests are meant to represent their construct, in practice it cannot be assumed that tests are construct-representative until validation has been carried out. This is due to the two threats to construct validity. One is construct-underrepresentation, occurring when tests fail to measure part of the elements specified in the test construct. The second is construct-irrelevant variance, which happens when the test is too broad and other factors such as background knowledge, test methods, and test-wise strategies contribute to success in test performance.

According to Messick (1995), construct validity is a superordinate term, which concerns not only evidence of the theoretical construct tests are designed to tap into but evidence related to content relevance, criterion-relatedness, and the social consequences of test use. Construct validity concerns not only the construct underlying the test itself but also the interpretations and meaning of test scores. The construct underlying tests is no longer viewed as only a component of language ability theoretically indicated but also the cognitive processes that individuals demonstrate during task performance. A set of construct indicators, which necessarily explain the construct underlying test tasks, thus includes cognitive processes, strategies, and linguistic and non-linguistic knowledge applied during task performance (Messick, 1995).

Messick's conceptualization of test validity has been widely accepted in language assessment and has triggered changes in test validation. However, it has been perceived as being rather abstract and lacking clear guidance for test validation in practice (Bachman, 2005; Davies, 2012). To address these gaps, several frameworks for test validation have been introduced. These include three influential frameworks, i.e. Kane (1992), Bachman (2005), and Weir (2005), each of which is briefly reviewed in the following paragraphs.

Kane's (1992) argument-based approach to test validity

One framework which has been widely acknowledged in the language testing literature is Kane's (1992) argument-based approach to test validity (see Chapelle, 2012; Xi, 2008). In agreement with Messick (1989), Kane (1992) proposed that test validity is related to the interpretations and decisions made on test scores rather than the quality of tests themselves. Kane described test validation in relation to two kinds of argument building: 1) an interpretive argument and 2) a validity argument. The interpretive argument states the proposed interpretations of test scores. The interpretive argument, as illustrated later by Kane, Crooks, and Cohen (1999), involves a series of links from test scores (performance) to score interpretations (see Figure 2.1). First is a scoring link, i.e. the link from test performances to observed scores. The second link is generalization which links the observed scores obtained from test tasks to universe scores obtained from other tasks similar to the assessment tasks. The third link is extrapolation, i.e. the link from universe scores to target scores. This link leads to the interpretation of test scores, i.e. what test-takers can do in the target language use domain.

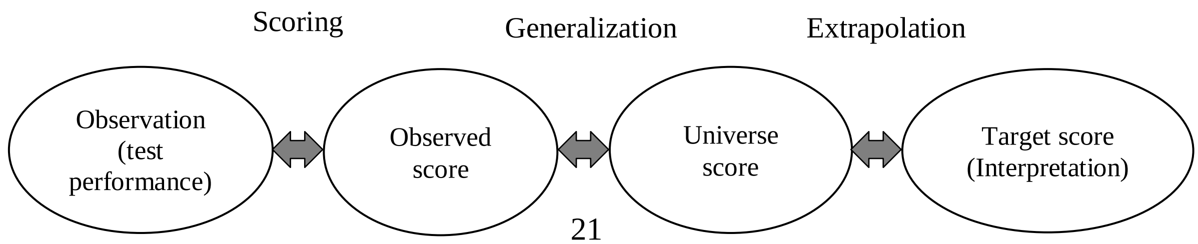


Figure 2.1: Links in interpretative argument by Kane et al. (1999, p.9)

The second type of argument, as Kane describes, is the validity argument which critically evaluates the interpretative argument, i.e. the inferential links proposed in the first stage. Three basic criteria are recommended to evaluate the argumentation: 1) the clarity of the inference network, i.e. whether enough and clear details are provided to back up and warrant the links, 2) the coherence of the argument consisting of logical and convincing links from test performance to the interpretation of test scores, and 3) the plausibility of inferences and assumptions.

The credibility of these links, as Kane et al. (1999) contend, rests on several types of evidence. The scoring link, in particular, relies on at least two types of evidence: 1) evidence showing that the scoring criteria are reasonable and employed correctly and 2) evidence indicating the performance occurs under conditions relevant to those of the target situation where the test scores are proposed to be generalized to. The generalization link, which assumes that observed scores represent the performances in the target situations, relies on reliability studies of the measure used. The credibility of extrapolation depends, for example, on the similarity of the language abilities inferred from the universe scores to those required in the target situations.

Despite providing a logical set of procedures to validate tests, Kane's framework has been criticized for focusing only on score interpretations. As Bachman (2005) contends, it does not link the score interpretations to the use (decision) and the social consequences of the scores being used. Although Kane (2006) attempted to address the use of test scores, McNamara (2006) comments that it does not explicitly explain what, as Messick (1989) suggests, the social consequences of the tests are. To fill the gaps in

Kane's framework, Bachman (2005) proposed his assessment use argument (AUA) for test validation.

Bachman's (2005) AUA for test validation

Bachman (2005) explained that the essence of an AUA is the link between the overall arguments to the social consequences of test scores. In the AUA framework, test validation is described in relation to two types of arguments: 1) an assessment utilization argument, linking test performance to a decision, and 2) an assessment validity argument, which justifies an assessment utilization argument. The assessment utilization argument is an articulation of claims (proposed interpretations of test use), warrants (statements to support the claims), and rebuttals (statements to reject the claims).

Four types of warrants, as Bachman indicates, are necessary to support claims: 1) relevance, 2) utility, 3) intended consequences, and 4) sufficiency. Relevance concerns the extent to which the ability assessed by the test represents that needed in the target language use (TLU) domain, and the relevance of the task characteristics to those in the TLU. Utility deals with the extent to which interpretations made on the basis of test performance are useful in making the intended decision. Intended consequences concern the extent to which the intended decision will benefit the individuals involved in the use of test scores. Sufficiency is considering whether the assessment provides sufficient information for decision making.

The assessment validity argument is an evaluation of the claims being made by providing different types of evidence to back up the claims or decisions made on test scores. Such evidence includes, for example, evidence on construct validity and task authenticity which warrants task relevance, or scoring reliabilities which warrant utility.

This review of Kane's and Bachman's frameworks has indicated that the scope and nature of test validation defined by more recent conceptualizations of test validity (i.e. Messick, 1989) differs considerably from those described in the earlier views. Test validity in these frameworks is no longer regarded primarily as the qualities residing in tests but as the logical links between performances observed in test situations to score interpretations and use. The credibility of such links (which indicate test validity) requires different sources of evidence, including, for example, context relevance, scoring reliabilities, and criterion-related validity. Although Kane's and Bachman's frameworks have emphasized that different types of evidence are required to justify test validity, Weir (2005) claimed that such evidence may only support propositions of score interpretations and use made by testers and led by assessment tasks. As such evidence is only gathered after test events, it may not be sufficient to support score interpretations specified at the initial stage of test design (Weir, 2005). To address this gap, Weir (2005) put forward his socio-cognitive framework for test validation.

Weir's (2005) socio-cognitive framework

In line with Messick (1989), Weir (2005) proposed that test validity involves several types of evidence to support any claims made on the basis of test scores; no single type of validity is considered superior to another. In his framework, Weir (2005) presents two main categories of validity evidence – *a priori* and *a posteriori* evidence – each of which comes along with subcategories. Prior to actual test administration, Weir requests evidence of theory-based validity and context validity. Theory-based validity, which is later referred to as cognitive validity (see Taylor & Geranpayeh, 2011; Field, 2013), concerns the theoretical construct of ability, knowledge, and processing expected to be captured by test tasks. Context validity concerns the extent to which a test task is representative of real-world tasks, especially in terms of its linguistic and content demands as well as the conditions under which the task is performed. The *posteriori* or after-test event evidence comprises scoring validity, criterion-related validity, and consequential validity. Scoring validity is the extent to which test results are stable over time and across different measures. It accounts for the degree to which examination marks are free from errors of measurement and the extent to which they can be depended on when decisions are made about test-takers. Criterion-related validity involves the correlations of test scores to other measurements of performance. Consequential validity discusses how the interpretations and use of a test impact on the individuals involved in an assessment (e.g., test-takers, teachers, and parents) and on society as a whole. As emphasized by Weir (2005), all these validity sources are needed to support a discussion of test validity.

While Kane's and Bachman's validation frameworks rely principally on the evidence collected after test events, Weir (2005) explicitly indicated that test validity is not only a matter of *a posteriori* data analysis (to support the logical inferences drawn from test performance to score interpretations and use). Test validity, according to Weir, also concerns a before-test event investigation or an investigation of the abilities needed in the TLU domain and the theoretical concepts of the language abilities test tasks have the potential to assess. This is in order to inform test design and development (Weir, 2005). After the test event, the theoretical construct of language abilities as preliminarily indicated are compared to those actually assessed by the test tasks in order to determine the extent to which tasks can actually capture the abilities intended to be tested from the start.

According to Weir (2005), having clear test constructs from the initial stage is essential. Testers, as Weir emphasizes, ‘can never escape from the need to define what is being measured, just as we are obliged to investigate how adequate a test is in operation’ (p.18). This view has, for example, been supported by an empirically-based study of test validity in the context of integrated tests, i.e. Frost, Elder, and Wigglesworth (2011). Specifically, this study has indicated that test constructs or descriptions of what integrated tasks measure are crucial and have to be provided at the start of test validation. This is because test constructs are used as the basis for the interpretations of test-takers’ abilities (what test-takers can do). When test constructs are unclear, it is therefore difficult to provide explicit and defensible links between test performance and the interpretations of test scores and use.

Particularly in the case of integrated-test tasks, which are the focus of this study, previous research (e.g., Cumming et al., 2006; Frost et al., 2011) has shown that it is difficult to establish explicit and logical links from test performance to score interpretations and use. This is partly because this task type involves a combination of language skills in task performance and this makes the construct underlying these tasks complex. In addition, the nature of the interaction and the contribution of each skill involved in task performance are not yet fully understood. In this regard, and in alignment with the more recent conceptualizations of test validity, specifically Messick (1989) and Weir (2005), this study set out to investigate the listening construct underlying integrated-listening tasks, i.e. what listening abilities assessed by listening-to-summarize tasks and what test-takers' perceptions of tasks and of listening task difficulty. Findings from this investigation will throw light on what Weir (2005) terms 'theory-based or cognitive validity' and context relevance (as perceived by test-takers) in the socio-cognitive framework for test validation. Since a big concern over the use of integrated-test tasks is that the theoretical construct of this task type is not fully understood or developed, Weir's (2005) approach, which emphasizes that test constructs have to be clearly defined at the beginning of test design and development and which explains clearly how such construct can be established, offers a useful and suitable framework for this study.

In the next section (2.3), a description of integrated-skills assessments, definitions and the use of integrated-test tasks in language assessment will be given, to clarify the nature of this form of assessment and the need for research on it in relation to test validity. Previous studies on this task type are also reviewed in order to determine what needs to be done in order to investigate the abilities assessed by this task type.

2.3 Integrated-skills assessments

The assessment of second language performance was traditionally largely in favour of discrete-point and indirect testing until the 1970s. With the increasing implementation of communicative language teaching approaches, however, the assessment of language proficiency started shifting more to assessing language skill performance such as assessing listening, speaking, reading, and writing (Weir, 1990).

Despite this shift in focus of language assessment, concerns related to skills assessment have been expressed, partly because skills assessment has been carried out in isolation. For example, Ferris and Tagg (1996) and Frost et al. (2011) have questioned the authenticity of the language test tasks often used, given that human communication is a dynamic, complex and interdependent system. East (2012) in addition argues that the assessment of the four language skills in isolation, though communicative in orientation, does not reflect the interactive nature of language used in real-life contexts. McNamara (2000) mentions that assessing one language skill such as listening separately from speaking could have validity issues as in most oral communication these two skills are typically applied together. Therefore, integrated-skills assessments have been introduced and implemented, including in large-scale tests such as the TOEFL iBT, aiming to address the concern that the assessment of individual skills in isolation may be inadequate to represent the ability to use language in real-life situations.

2.3.1 Defining integrated-test tasks

In order to understand how integrated-test tasks have been used in language assessment, it is essential to look at how these tasks are defined. Lewkowicz (1997) defined integrated tests as tests "where the input that has been provided forms the basis for the response(s) to be generated by test-takers" (p.121). However, this definition is very broad in the sense that it could equally apply to independent speaking or writing tasks that require test-takers to simply discuss a topic provided.

More detailed definitions have been given in research specifically focussing on integrated-test tasks (e.g., Asencion, 2004; Brown, Iwashita, & McNamara, 2005; Cumming et al., 2006; Frost et al., 2011; Plakans, 2009; Weigle, 2004). Brown et al.

(2005), for instance, defined integrated-test tasks as tasks that require test-takers to process input material in any form in order to integrate the information from this source into task performance. This input processing, as Brown et al. (2005) argue, makes integrated-test tasks more complex and more cognitively demanding than independent-test tasks, such as writing-only or speaking-only tasks, which only require the test-takers to exclusively draw on their own knowledge or ideas to respond to questions or prompts. Similar to Brown et al. (2005), Cumming et al. (2006) refer to integrated-test tasks as tasks that combine language skills – reading, writing, listening, and speaking – in various ways. They explain that integrated-test tasks differ from independent-test tasks in that they require references to source materials while independent tasks do not.

What is emphasized in the definitions given in these previous studies is the provision of input materials, input processing, and the combination of language skills. Input, prompts, or stimulus materials – the terms interchangeably used in previous research on integrated-test tasks – comprise a source text which test-takers have to comprehend by reading and/or listening and on the basis of which produce task output. Among the three components, a source text seems to be the most important feature of integrated-test tasks because it stimulates language processing and the transferring of knowledge from the input into task performance in either oral or written form or both.

Drawing on the above definitions, this study describes integrated-test tasks in relation to three important task components; namely, a source text, language processing and transferring, and task production (see Figure 2.2).

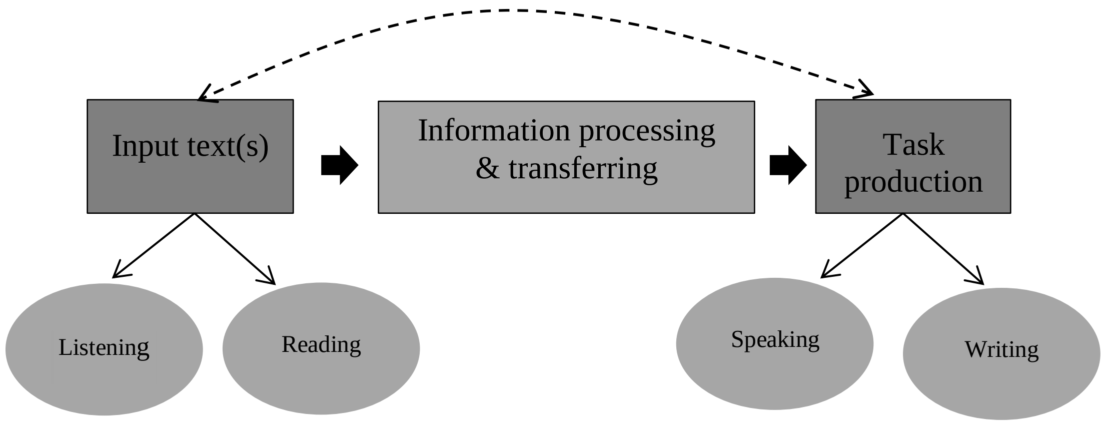


Figure 2.2: The components of integrated-test tasks

Integrated-test tasks require language processing that involves references to at least one source text in task production. Therefore input material is provided in order to stimulate language processing and information transferring. The tasks begin with requiring test-takers to read or listen to input text or both (Asención, 2004; Asención Delaney, 2008; Plakans, 2008). Then based on their understanding of the input text, test-takers are required to produce task output either by speaking or writing or both, depending on the test objectives, to demonstrate their language ability. During task production, test-takers may revisit task (written) input or checking their reading/listening notes.

2.3.2 Previous research on integrated-test tasks

A body of research has been carried out along the implementation of integrated test tasks. One line of research has been to compare the construct underlying independent-skill tasks (e.g., an independent-writing task) with integrated tasks (e.g., a reading-to-write task). Another strand of research has investigated the performance aspects that are associated with proficiency levels. This section reviews previous research in these two areas to point out gaps in the previous research and provide justification for this study.

Comparing the construct underlying independent- and integrated-test tasks

This review of the research on integrated-test tasks has suggested that the early research set out with the aim of describing the construct underlying integrated-test tasks or abilities actually measured by this task type. These studies have compared test-takers' performances on independent-test tasks, e.g., independent-writing, to performances on integrated-test tasks, e.g., reading-writing, in order to study whether the two task types measure the same construct. Two different sources of data have been used to achieve this aim (see Table 2.1). One involves comparing performance scores obtained from independent-skill tasks to those on integrated-test tasks. The other involves comparing test-takers' task processing behaviours on the two different task types.

The findings from the studies presented in Table 2.1 provide inconclusive evidence in terms of the construct underlying integrated-test tasks, particularly where the quantitative studies are concerned. Both Lee (2006) and Gebril (2010) found that the scores on integrated-test tasks were highly correlated with those from independent-skill tasks. These studies, thus, concluded that the two task types measure a similar construct. However, Asención Delaney (2008) and Sawaki et al. (2009) found that the scores on independent-skill tasks were not significantly related to those on integrated-test tasks. These studies suggested that these two task types are different in terms of what they assess. Potentially, these conflicting findings are due to the different methods of statistical analysis applied. G-theory, which was used in Lee (2006) and Gebril (2010), and CFA, used in Sawaki et al. (2009), took into account the effects of other factors such as raters and tasks on test scores in the analysis whereas the correlation analysis used in Asención Delaney (2008) did not. Other factors that might also add to the inconclusive results could be differences in test-taker characteristics, tasks, and raters, all of which, as noted by these researchers, can affect the test results.

Researchers (year)	Task types[1]	Research procedures	Research Data	Results	
				Similar	Different
Lee (2006)	• Independent speaking • Listening-speaking • Reading-speaking	G-theory[2]	Task performance scores	✓	
Asención Delaney (2008)	• Reading-only • Writing-only • Reading-to-write	Correlation analysis	Task performance scores		✓
Plakans (2008)	• Writing-only • Reading-to-write	Talk-aloud in writing sessions	Test-takers' cognitive processes		✓
Sawaki, Stricker, and Oranjie (2009)	• Independent speaking • Integrated speaking	Confirmatory factor analysis (CFA)[3]	Task performance scores		✓
	• Independent writing • Integrated writing				✓
Gebril (2010)	• Writing-only • Reading-to-write	G-theory	Test-takers' cognitive processes	✓	

Table 2.1: Previous research comparing the construct underlying integrated-test tasks and

[1]Task names in this column were adopted from the research where they appeared. Although they have been named slightly differently, they are categorized as either independent-skill or integrated-test tasks, depending on the number of skills involved in task performance.

[2]G-theory is a methodology used to examine the generalizability of test scores when there is more than one major facet involved in assessment, for example, in speaking assessment when tasks and raters can be major sources of score variability (Lee, 2006).

[3]A confirmatory factor analysis (CFA) is a statistical analysis used to analyze test constructs. It allows testing researchers to test whether test data (scores) fit a theoretical construct (proposed/hypothesized before test events) by indicating relationships between test items and the extent to which these items tap into the same construct (Green, 2013). In Sawaki et al. (2009), it was used to conduct a fine-grained analysis of individual items and the relationships of the integrated items to the test sections.

independent-skill tasks

Results of the qualitative studies which looked into test-takers' cognitive processes in task performance are congruent with Asención Delaney's (2008) quantitative study's findings. These studies (e.g., Plakans, 2008) showed that the mental processes performed to complete independent-skill tasks were different from those performed during integrated-test tasks. Plakans (2008), for example, found that a writing-only test task required more effort planning content while the reading-to-write task required more thinking for task interpretation. The reading-to-write task, however, was likely to demand a more interactive process in that the writers had to read the source text, interact with it, and formulate ideas and opinions on the topic. The writers, as Plakans (2008) explained, engaged more in meaning-making and making inferences in the reading-to-write task. Based on the differences in the processing behaviours in both tasks, the researcher suggested that the construct underlying integrated-test tasks is different from that of independent-skill tasks.

Comparing performance characteristics between task types and performance levels

Another line of research on integrated-test tasks, as presented in Table 2.2, compares the characteristics of performance between task types and performance levels: high, average, and low. Discourse-based analysis has been employed in these studies. The results show that performance features in independent- skill and integrated-test tasks are generally different. Brown et al. (2005), for example, found that the performances on independent-speaking test tasks were different from those on integrated-speaking test tasks in all four features investigated, namely linguistic resources, phonology, fluency, and content.

Researchers (year)	**Task types**	**Research procedures**	**Features analysed**	**Results: Features which are different when compared between:**	
				Task types	**Performance levels**
Brown et al. (2005)	• Independent speaking task • Integrated listening-speaking task • Integrated reading-speaking task	Discourse-analytic approach	• Linguistic resources • Phonology • Fluency • Content	✓ ✓ ✓ ✓	N/A
Cumming et al. (2006)	• Independent writing • Integrated-reading-writing • Integrated listening-writing	Discourse-analytic approach	• Lexical complexity • Syntactic complexity • Grammatical accuracy • Argument structure • Orientations to evidence, • Verbatim uses of source test	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓
Gebril and Plakans (2013)	• Integrated reading-writing	Discourse-analytic approach	• Fluency • Lexical sophistication • Syntactic complexity • Grammatical accuracy • Verbatim source use	N/A	✓ ✓ ✓ ✓ ✓

Table 2.2: Previous research comparing performance characterisitcs between task types and performance levels

The linguistic features (vocabulary, grammar, and schematic structure) and content (ideas shown in the responses) were more complicated in integrated-speaking task performances than those in independent-speaking tasks. This could be because, as hypothesized by the judges in this study, the input materials provided test-takers with more language input to rely on when completing the tasks. However, more difficulty in pronouncing key words and more disfluency were observed in integrated-speaking-test tasks than in independent-speaking test tasks. The researchers speculated that this might be because of lexical difficulties caused by the input in the integrated-speaking tasks.

Cumming et al. (2006) similarly found that the characteristics of the written performance of independent- and those of integrated-writing tasks differed. That is, in the integrated-writing tasks, test-takers used more complicated words with a wider range, more and longer clauses, and more verbatim phrases than they did in the independent-writing tasks. The test-takers also engaged more in paraphrasing and summarizing than stating personal knowledge. The researchers argued that this is because of the different nature of the two task types and of the prompts provided. The independent-writing tasks required test-takers to form coherent argumentative essays based on personal knowledge and experience, whereas the integrated-writing test tasks required complex cognitive and language abilities for understanding input materials and producing essays that demonstrate appropriate and meaningful uses of source materials. The integrated-writing tasks in Cumming et al. (2006) were therefore considered to be more cognitively demanding than the independent-writing tasks.

Although Brown et al. (2005) and Cumming et al. (2006) agreed that integrated-test tasks can be used to assess productive skills, they hypothesized that the comprehension part of the integrated-task performance affects task production. Brown et al. (2005), in particular, thought that the lower fluency and greater disfluency in the integrated-speaking performances might be the result of low text comprehension. To gain more insight into this issue, they proposed looking into test-takers' task processing behaviours. This source of data, as they emphasized, could provide further evidence to define the construct of integrated-test tasks.

Other studies investigating performance features point to some associations between performance levels and discourse features. In integrated reading-writing test tasks, Gebril and Plakans (2013) found that the features of fluency, grammatical accuracy, over-all source use, and indirect source use differed across three proficiency levels. Fluency, however, was the only feature that distinguished all three levels from each other. Grammatical accuracy and the two-source use features were significantly different between the lowest level and the other levels, but not between the upper two levels. Since the accurate and appropriate use of source materials involved reading ability and this was the feature that separated the high from the low scorers, these researchers suggested that reading proficiency and knowledge about discourse synthesis should be recognized to be part of the construct of integrated reading-writing tasks.

To sum up, previous studies comparing task performance characteristics between integrated-skill and independent-test tasks have shown that the abilities captured by these two task types, although some overlapped, differed in several respects. This is partly because of the different nature of the tasks and the effects of comprehension ability. In addition, when focusing on integrated tasks only, the studies showed that performance characteristics were associated with different performance levels; high-scorers in integrated listening-speaking tasks were able to use more difficult words and complex structures and provide more accurate content.

In conclusion, the review of research on integrated-test tasks indicates conflicting views on the construct underlying this task type. Some quantitative studies (Gebril, 2010; Lee, 2006) indicate that the performance scores on integrated-test tasks are highly related to those on independent-skill tasks, suggesting that these two task types measure a similar (or the same) construct. Based on this, integrated-test tasks have been used to assess productive skills (speaking and writing). On the other hand, the studies investigating test-takers' processes (e.g., Asencion, 2004; Plakans, 2008) found that these two task types are different in terms of what they assess. The ability to comprehend a source text, either in a spoken or written form, requires a certain amount of text comprehension ability (Brown et al., 2005; Cumming et al., 2006). This ability does not only distinguish performances on integrated-skill tasks from independent-test tasks (see Brown et al., 2005; Cumming et al., 2006) but is also associated with different performance levels (see Cumming et al., 2006; Gebril & Plakans, 2013). These studies support the idea that the constructs underlying these two task types are different. Brown et al. (2005) and Cumming et al. (2006), in addition, found that features such as lexical complexity and grammatical and schematic structures differed considerably between performances on integrated-skill and independent-test tasks. These studies hypothesized that comprehension played a role and had an impact on integrated-test task performances such as on difficulty in pronunciation and disfluency in speaking. These studies, as a result, suggest looking into test-takers' cognitive processes in order to better define the construct underlying the tasks.

The concerns over the use of integrated-test tasks, as so far presented, are related to the construct underlying the tasks, and, in particular, the fact that it is not clear what abilities are actually assessed by this task type. This is especially true for the receptive skills (listening and reading) involved in task performance. It should be noted that, as evidenced in Tables 2.1 and 2.2, existing research has primarily looked into integrated-test tasks with reading input. Given that the these researchers have urged for the recognition of receptive skills as part of the construct of integrated tasks, the need for more research exploring the role of listening in the much less researched integrated tasks with listening input is clearly substantiated. Therefore, this study set out to describe the listening construct of integrated-listening tasks, i.e., listening-to-summarize tasks.

As described in section 2.2 , what ‘test construct’ constitutes has been described in different manners, depending on how test validity is conceptualized. Traditionally, test constructs have been described by relying on the theoretical concept of abilities, and construct validity has investigated whether the target construct is measured by the test used. This has been done mainly through statistical analyses. In more recent conceptualizations of test validity, construct validity is no longer an indication of whether a theoretical construct is assessed by tests but a justification of the interpretations of test scores and use. Test constructs, according to these recent views, are described in association with the context abilities are performed in and by looking into test-takers’ task processing behaviours. Following this recent view of test validity, this study conceptualizes the construct underlying test tasks as the cognitive processes and strategies test-takers engage in while performing the tasks. Data on test-takers’ cognitive listening processing will thus be gathered in order to describe the listening construct of listening-to-summarize tasks. The next section reviews theoretical descriptions of cognitive processes involved in language comprehension in general and then specifically in listening comprehension processing. This is to guide the analysis of listeners' processing behaviours in this study.

2.4 Frameworks for language comprehension processing

According to psycholinguists, such as Anderson (1985), Call (1985), Færch and Kasper (1986), and Garrod (1986), text comprehension is the product of several cognitive subsystems and metacognitive strategies working in a parallel and interactive manner. As its end product, the comprehension process provides a mental representation of a text, which is a network of interrelated propositions (Kintsch & van Dijk, 1978).

To understand how comprehension is reached through processing, four cognitive processing models are reviewed in this section. First, Anderson's (1985) cognitive framework for language comprehension which applies to both listening and reading comprehension is reviewed (2.4.1). Although developed with reference to L1 comprehension, Anderson's model is frequently quoted in the L2 literature.

Then two models are described (2.4.2-2.4.3), Rost's (2011) model of listening processing and Vandergrift and Goh's (2012) cognitive model of L2 listening comprehension, both specifically designed to explain listening comprehension processes in relation to L1 and L2. This is in order to understand how L2 listening processing is different from L1 listening processing and what contributes to success in L2 listening. It is important to note that these two models are reviewed as complementing Field's (2013) framework, which is presented in section 2.4.4 and which was used to analyse this study's pilot study data. Since this study is conducted in the listening testing context, Field's (2013) cognitive processing framework for listening, introduced in this context, was considered relevant and used to analyse listening processing behaviours in this study's pilot study. However, as will be shown in the pilot study (see 3.5), some aspects of processing behaviours contributing to successful task performance were not explained in Field (2013). Therefore, the review extends to these further models used in second language acquisition (SLA) research.

The four models are then compared (2.4.5) and this leads to a description of the role of strategies used in L2 processing (2.4.6). The listening model used to analyze listening processing behaviors in this study is then presented (2.4.7).

2.4.1 Anderson's (1985) cognitive framework for language comprehension

One of the influential models used to explain how language comprehension takes place is Anderson's (1985) cognitive framework for language comprehension. Anderson (1985) distinguishes between three stages of mental processing in the comprehension of aural and written texts, namely perceptual processing, parsing, and utilization, each of which will be explained in the following paragraphs.

Perceptual processing

Perceptual processing, which is the lowest stage of language processing, entails the processes by which an acoustic or written message is originally decoded. According to Anderson (1985), perception involves registering information arriving at one's eyes and ears. Generally, there is a large amount of information coming at the same time and this information is not retained unless it is registered and transferred to short term memory (Anderson, 1985). In most cases, unregistered information decays within a second (Loftus & Loftus, 1976).

Specifically in the case of speech or auditory texts, Anderson (1985) explains that perception involves detecting phonemes in continuous speech which are the basis of utterances and grouping them into words. When the sounds are perceived and delivered to short-term memory, some initial analyses of the language code and encoding process begin. The sounds will be converted into phonological forms and later grouped into categories to create meaningful representations, i.e. words. During this process, attention may be selectively directed to aspects of the context that will be useful for text decoding such as pauses and acoustic emphases (O'Malley & Chamot, 1990).

Parsing

Parsing, according to Anderson (1985), is the process by which language users segment a text into chunks of information which are meaningful to them. It is a more automated and more precise stage of input processing that takes place when the users relate the sounds/words perceived to their knowledge. A basic unit obtained from this stage of processing is propositions or chunks of information (Anderson, 1985; Kintsch & van Dijk, 1978).

Anderson (1985) explains that parsing relies on two types of knowledge, i.e. syntactic and semantic knowledge. That is, after words are recognized, processing the language for meaning requires a partial syntactic mapping of those words onto a grammatical structure. A basic form of syntactic cue that can guide parsing is word order. With the knowledge of English word order, the users can tell that the two sentences, “the dog bit the cat" and "the cat bit the dog" have different meanings although they contain the same words. Another syntactic cue that is beneficial to parsing is grammatical or function words, such as 'the….of' and 'who', because they indicate chunks of information in connected speech. Comparing the following sentences, a) "the boy whom the girl liked was sick", and b) "the boy the girl liked was sick" (Anderson, 1985, p.394), Anderson (1985) explains that the first sentence (a) is more easily parsed because the relative pronoun 'whom' signals a chunk of information. In addition to syntactic structures, the listener can use semantic cues to guide parsing. The listener can understand that 'Jane fruit eat' means 'Jane eats fruit' although the sentence does not correspond to the syntax of English (Andersion, 1985, p.394). This is because sometimes people just rely on plausible semantic interpretations of words in a sentence, not on syntactic structures (Anderson, 1985).

Utilization

Anderson (1985) describes utilization as the process of combining parsed propositions with the individual's external knowledge in order to comprehend the entire meaning of a text. Text comprehension involves semantic processing at two levels, i.e. microstructure and macrostructure (Kintsch & van Dijk, 1978). Microstructure processing is a local level of semantic processing which provides a conceptual link between individual propositions. Macrostructure processing connects related propositions to the theme of a text to create its discourse meaning. During macrostructure processing, propositions might be deleted if neither direct nor indirect interpretation of the propositions is made in relation to the topic. Some propositions might be substituted by more general propositions to encompass an immediate superset or a global fact.

Since texts are not a list of unrelated propositions but a coherent structure which, in most cases, is not explicitly indicated, a language user has to assign a structure to the texts in different manners, including making inferences and elaborations. A number of inferences and links between linguistic information and world knowledge are made to understand the discourse meaning of texts. Anderson (1985) distinguished two categories of inferences , i.e. backward inferences and forward inferences. Backward inferences connect the current sentence to prior sentences or to background knowledge to identify how parts of the text fit together. One important aspect of backward inferences is recognizing when an expression in a sentence refers to something already stated in the previous sentences by the use of the definite article “the” or pronominal reference. When there are multiple possible candidates for the referent of a pronoun, syntactic and semantic cues will be called on to assist the selection of a referent. Forward inferences, on the other hand, are the use of linguistic information from previous parts of the text to anticipate incoming information or future consequences on the text.

Elaboration, which occurs at both micro and macro levels of semantic processing, is the use of prior knowledge to assist text recalls and comprehension (Kintsch & van Dijk, 1978). Three ways of elaboration that have been found to facilitate comprehension are 1) linking textual information to one’s own world knowledge, 2) connecting new information to something meaningful at a personal level, and 3) asking questions about the text or anticipating possible extension of the information (O'Malley, Chamot, & Kupper, 1989). Elaboration can also be used to bridge gaps in inferences and to infer the meaning of unfamiliar words (O'Malley et al., 1989).

Although Anderson (1985) indicates that his three-stage comprehension framework is applicable to both reading and listening comprehension processing, he does not explain what is specific to listening or reading comprehension processing respectively. Regarding that comprehension processing must be related to the form of texts (i.e., aural and written forms) the users process, this model may overlook some processing behaviours specific to text modality. In this regard, frameworks developed specifically for listening comprehension are reviewed in the next sections in order to provide a more comprehensive view of how listening processing works and what L2 listening processing in particular entails.

2.4.2 Rost's (2011) listening processing

Rost's (2011) model, as recommended by Weir (2005), is one of the well-informed models for teaching and researching listening. Considering listening as involving overlapping types of processing, Rost (2011) proposes four categories of listening processing, i.e. neurological processing, linguistic processing, semantic processing, and pragmatic processing.

Neurological processing

Rost (2011) explains that neurological processing, the start of listening processing, is when sounds are heard and transmitted to short-term memory. Listening is about continually gathering incoming and perceived sounds that can stay in a memory buffer for a few seconds after being heard. The acoustic sounds which are perceived will be collected in a mental package and used to identify what is heard, what is to be heard next, and what has just been heard.

Linguistic processing

Linguistic processing, which is fundamental for text comprehension, involves processing at two levels, i.e. word recognition and parsing. After acoustic sounds have been perceived and registered in short-term memory, this processing goes on to identify words which are the basis of speech units. Word recognition, as Rost (2011) describes, has two different patterns, i.e. top-down and bottom-up processing. Top-down processing makes use of other types of knowledge such as contextual and world knowledge to put together the acoustic information (Rost, 2011). Bottom-up processing, on the other hand, involves grouping perceived acoustic sounds into words or phrases. No matter which word recognition processing pattern is applied (and in most situations both are utilized), effective word recognition entails two synchronous tasks, i.e. identifying words and lexical phrases and activating knowledge associated with those words and phrases (Rost, 2011).

Parsing is when recognized words are mapped onto a language's grammatical structure. Spoken language parsing entails assigning grammatical categories (e.g., word forms) to the recognized words or phrases, and creating a form-meaning relation between these categories. At this stage of processing, meaningful chunks of information will be identified and a propositional model of the incoming text will be constructed. To facilitate the process, the listener's syntactic and morphological knowledge of word form, word order, and subject-verb agreement, is applied (Rost, 2011).

In addition to syntactic knowledge, parsing is facilitated by language familiarity, including the listener's familiarity with common sequences of formulaic language (a continuous or discontinuous string of words) and other sources of knowledge stored in and retrieved from long-term memory at the time of use or interpretation. Processing familiar strings of words is, thus, easier than unfamiliar language. The knowledge of context-appropriate prosody with the ability to attend to pitch levels also benefits parsing. Different pitches indicate pause units which show newness, separateness, connectedness, incompletion, or completion of the incoming information so that the listener can recognize phrase and sentence boundaries (Rost, 2011).

Although word recognition is an important contributor to listening comprehension, under some conditions such as limited lexical knowledge, noise, or other perceptual stress situations, word recognition may not be successful. However, comprehension processing can often continue successfully even if not every word in a speech stream is recognized (Rost, 2011). This is because the listener can make inferences about the meaning of an utterance through other sources such as topical and pragmatic knowledge (Rost, 2011). Successful listeners, however, must be able to tolerate ambiguity, and wait for later utterances to decide what was intended before.

Semantic processing

Listening comprehension, according to Rost (2011), is a process of semantic mapping and updating in listeners' memory, which is facilitated by the amount of schematic structure and social common ground shared with the speaker. Comprehension occurs when listeners can relate the incoming text to concepts in their memory and their world knowledge. This can be achieved by connecting different parts of the utterances together to create a mental representation of the text and develop a figurative map in which new information and concepts will fit.

What is necessary during the process of semantic mapping and meaning construction, as Rost (2011) explains, is inferencing. Conventional language knowledge, such as an understanding of cohesive devices, assists semantic mapping. However, only this knowledge may not be sufficient for full text understanding (Rost, 2011). Since in most cases speakers do not provide clear links between the various bits of utterances or they do not explicitly state their intentions, listeners have to infer the links and the intended meaning of a message by using their background knowledge. Semantic processing, thus, involves not only finding coherence within the language used but also inferring what is left unsaid by the speakers by relying on logic and listeners' real-world knowledge.

Rost (2011) explains that semantic processing occasionally involves compensation strategies. This is because comprehension relies on cognitive processes activated in short-term and long-term memory. As the processing capacity of both types of memory is naturally limited, semantic processing is obstructed from time to time. Problems that will probably occur during semantic processing are, for example, 1) it is not clear to the listener what the speaker is saying and what specific expressions the speaker is using, 2) the information the speaker gives is incomplete to the listener, 3) a familiar word is used in an unfamiliar way, and 4) unknown words are frequently used. To compensate for comprehension gaps resulting from these issues, strategies are employed. Commonly used strategies, as Rost (2011) indicates, are 1) skipping or omitting a part or a block of the text from processing, 2) using a superordinate concept or constructing a less precise meaning for an unclear word or concept, 3) filtering or compressing a longer message or set of propositions into a more concise one, 4) maintaining an incomplete proposition in memory, 5) waiting until clarification can be obtained, and 6) substituting a word or a concept or a proposition for one that is not understood.

Pragmatic processing

Pragmatic competence in listening, as described by Rost (2011), includes the abilities to 1) understand the intentions and strategies speakers use to communicate their ideas, 2) use contextual information and knowledge of the social conventions of language use for text comprehension, and 3) elaborate the speaker input on a familiar context. The central aim of pragmatic processing is to derive and build the contextual meaning of utterances by integrating the interactional status and interpersonal relations between the speaker and the listener into the process. To understand the speaker's intentions, the listener might have to enrich the speaker's input in two ways: 1) inferring the speaker's emotion, which is generally implicit, and 2) elaborating the speaker's meaning by making semantic inferences on the concepts used by the speaker together with pragmatic inferences on world knowledge. According to Rost (2011), key pragmatic notions contributing to a listener's understanding of spoken language include 1) anchoring the utterance to a real-world situation or interpreting the utterance with respect to the physical context occurring in the real word, and 2) making references to the real-word context. To facilitate these processes, attention should be paid to what Rost calls the 'deictic elements' of an utterance, such as time adverbials and tenses, personal references, directional forces, demonstrative pointers, speaker's intentions and strategies, and the conversational implications of a message.

Despite the fact that Rost's (2011) model presents very clearly what cognitive listening processing entails, it relies mainly on insights into L1 listening. Although the model is occasionally extended to explain L2 listening processing, such as the use of compensation strategies to solve listening problems, it does not describe other strategies such as metacognitive strategies (goal setting, directed attention, and comprehension monitoring), which Goh (2002) has shown to play an important role in L2 processing. Although comprehension processes in L1 can be transferred to L2 listening, Færch and Kasper (1986) stress that they may require some adjustments because of limitations in linguistic knowledge and the lower automaticity of L2 processing. In the case of L2 reading in particular, Alderson (1984) indicated that L2 readers have to acquire some level of L2 competence before L1 reading ability is transferred to L2 reading. This also applies to L2 listening since researchers such as Buck (2001) and Cutler (2012) indicate that listening problems and gaps in understanding were found more in L2 listening than in L1. Bearing this in mind, Vandergrift and Goh's (2012) cognitive model which specifically aims to explain L2 listening comprehension processing is thus reviewed next.

2.4.3 Vandergrift and Goh's (2012) cognitive model of L2 listening comprehension

Vandergrift and Goh (2012) present a cognitive model to describe the cognitive processes and processing components involved in L2 listening comprehension. In their view, L2 listening consists of two main processing types, i.e. cognitive and metacognitive processing.

Drawing on Anderson's (1985) cognitive framework, Vandergrift and Goh (2012) describe cognitive L2 listening processing as an interactive process of speech perception, parsing, and utilization. These cognitive processes, as viewed by them, are bi-directional interactions between top-down and bottom-up processing which involve some degree of consciousness and are regulated by the listener's metacognition (Vandergrift & Goh, 2012). Metacognition is defined as "our ability to think about own thinking or cognition, and, by extension, to think about how we process information for a range of purposes and manage the way to do it" (Vandergrift & Goh, 2012, p. 83-34). Metacognition denotes a state of consciousness and can be observed via the strategies listeners use to manage comprehension and learning. Strategies are, thus, metacognitive due to the fact that they enable listeners/learners to purposefully change the way they use and learn language to improve their performance (Vandergrift & Goh, 2012).

According to Vandergrift and Goh (2012), the metacognitive processing which is activated to regulate cognitive processes and enhance listening comprehension entails planning for listening, monitoring comprehension, solving comprehension problems, and evaluating the approach and outcomes. Planning for listening refers to the processes by which the listener prepares to listen and establishes the necessary conditions to listen successfully. In order to plan for successful listening, the listener might, for instance, 1) bring to their consciousness their knowledge of the topic and relevant cultural knowledge, 2) anticipate words and ideas that they may hear, and 3) predict what they will hear based on the information brought to their consciousness and on relevant contextual information. Monitoring comprehension occurs while the listener is listening to a message. It is the process of evaluating comprehension and making necessary adjustments. The listener monitors their comprehension by, for example, 1) evaluating their understanding of texts, 2) checking if their predictions are consistent with the incoming text and their on-going interpretation matches world knowledge, 3) verifying their inaccurate predictions, 4) assessing the comprehension of desired information and necessary details, and 5) determining the effectiveness of their approach in understanding the text. Solving comprehension problems involves adjusting the approach to listening to the text and activating strategies to eliminate listening problems. Strategies include 1) inferencing the meaning of a chunk of a text that is not understood, 2) revising their predictions, and 3) adjusting their inferences to reflect new possibilities. Evaluating the listening approach and outcomes occurs when the listener reflects on their listening difficulties and on the success of their problem-solving efforts. Metacognitive processing, as noted by Vandergrift and Goh (2012), does not necessarily occur in a linear or circular process but in an interactive manner, depending on the listening context and the difficulty of the text listened to.

Similar to both Anderson (1985) and Rost (2011), Vandergrift and Goh's (2012) model describes listening comprehension as complex processing involving a number of cognitive processes including linguistic decoding, semantic mapping, and meaning construction, working interactively. While Rost (2011) focuses exclusively on cognitive processing with some references to the strategies used to compensate for gaps in comprehension, Vandergrift and Goh (2012) emphasize the role of metacognitive processing in listening, especially in L2 listening, where linguistic knowledge has not yet fully developed. Given the present study's interest in the listening construct, these models are essential. However, as the listening comprehension investigated in this study takes place in the context of test-task performance, it is also relevant to consider how language comprehension is described in the language testing literature. Therefore, the next section reviews Field's (2013) cognitive processing model formulated in the context of testing listening.

2.4.4 Field's (2013) cognitive framework for listening

Descriptions of language processing in language testing tend to relate to two levels of processing: lower- and higher-level processes. In assessing reading comprehension, for instance, Khalifa and Weir (2009) describe low-level processes in terms of local text understanding and high-level processes in relation to global text understanding. Text processing at the low or local level includes understanding word meaning and matching word-class to grammatical/syntactic structures to arrive at the basic meaning of propositions or the literal meaning of the text. At the high (global) level, readers go beyond the literal meaning of the text to infer further significance, build up a larger mental model, and identify text structure and purpose to form a representation of the text as a whole.

In a similar manner, within the context of assessing listening, comprehension processing, as described in Field's (2013) cognitive processing model, is classified into two levels, lower- and higher-level processes. According to Field (2013), processing at the two levels entails five types: 1) input decoding, 2) lexical search, 3) parsing, 4) meaning construction, and 5) discourse construction. The lower-level listening processes or linguistic processing involve the first three processes which take place when a message is being decoded. The higher-level processes are associated with meaning building and discourse constructing. Although the processes are presented in a linear order, Field (2013) emphasizes that it does not necessarily imply that one stage of processing waits upon another. Language processes, he notes, are often active in a parallel and interactive manner. The numbering is thus used only to distinguish between the processes. The following paragraphs describe the processes at each level.

Lower-level processes

The lower-level processes, according to Field (2013), start from recognizing acoustic input and develop to obtaining a phonological string by input decoding, a set of words by lexical searching, and an abstract proposition by parsing. Field explains that 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. This processing enables listeners to recognize strings of phonemes, some of which are marked as syllables or words. In the lexical search, listeners map sounds to spoken word forms. Based on their lexical knowledge, listeners determine word boundaries and identify words which are either content or function words in the connected speech. At the level of parsing, listeners separate units in the connected speech and construct propositions by applying their syntactic knowledge, an understanding of standard word order, and intonation group boundaries.

Higher-level processes

The higher-level processes involve two levels of processing; meaning and discourse construction (Field, 2013). Listeners start to construct the actual meaning of what they have heard by relating the propositions they obtained from the lower-level processing, which are context-independent, to their own schemata or the concepts of knowledge they have developed. At the level of meaning construction, it is the listeners' task to relate the propositions to the circumstances in which they were produced to obtain their full and relevant meaning. What the speaker said is often the raw meaning of the speaker's words and insufficient to convey the complete meaning of a text (Field, 2013). Listeners, therefore, have to supply information to comprehend what is said in a number of ways. One way to do this is using pragmatic knowledge to interpret the speaker's intentions. Listeners may also have to use contextual and semantic knowledge to relate the propositions to the context in which they occur. Listeners, in addition, may have to infer what the speaker left unsaid from what they have just heard or backtrack from what is being said to what has been said earlier.

Discourse construction is related to four processes that listeners apply to construct an understanding of a spoken text. Following Kintsch and van Dijk (1978), Field (2013) divides discourse construction into four processes: selecting, integrating, self-monitoring, and structure building. Selecting is deciding on the relevance of an incoming piece of information; for example, whether it is a repetition of a point made earlier or the central point of the topic being developed. On the basis of this decision, listeners may store, or ignore as irrelevant, the information being processed. Integrating is when listeners add a new piece of information to the discourse representation being developed. It involves recognizing conceptual links between the incoming information and the information processed before. Self-monitoring entails comparing whether a new piece of information is consistent with what has been processed before. If not, listeners consider whether the new judgment is correct, or question whether what they have understood and recalled earlier is correct. Structure building is prioritizing and organizing the information stored according to its importance and relevance.

Field (2013) stresses that the processes he describes are those operated by both L1 and proficient L2 listeners to understand the meaning of what has been said. As he cautions, these processes may not represent those activated by less skilled listeners. Successful listening, which means that the listeners have a clear concept of what the speaker intended to say, does not only depend upon linguistic processing (input decoding, lexical search, and syntactic parsing), but also higher-level processes (meaning and discourse construction). While the lower-level processes enable the listeners to produce propositions and to understand the literal meaning of the message being conveyed, the higher-level processes assist them to relate the incoming message to their existing knowledge and build a knowledge structure, resulting in complete understanding. According to Field, listening comprehension tests should therefore also tap into the higher-level processes.

2.4.5 Comparison of Field's (2013) listening comprehension model to Rost's (2011) and Vandergrift and Goh's (2012) models

In order to decide on a model of listening comprehension processing that can serve as a framework for analysing listening processing behaviours in this study, the models presented in sections 2.4.2 are compared. Specifically, Field's (2013) model, which has been introduced in the listening testing context is compared to Rost's (2011) and Vandergrift and Goh's (2012), proposed for language teaching and researching. This comparison indicates that there are similarities, differences, and overlap in their classifications of listening processes.

The consensus among these three models is that listening comprehension comprises cognitive processes that do not necessarily occur in a linear fashion but in an interactive manner and that listening processing does not rely only on linguistic knowledge but also on topical and world knowledge. Different terms appear to be used to refer to the classifications of cognitive listening processes. Rost (2011), in particular, divides listening processing into four categories according to the knowledge used during listening processing and refers to these categories as neurological, linguistic, semantic, and pragmatic processing. Based on the functions of processes, Field (2013), in a different manner, classifies cognitive listening processes into two main categories, i.e. lower-level and higher-level processes. This is in particular for the purpose of communicative language assessment, where higher-level cognitive processes are intended to be assessed. Despite these differences, Field's (2013) and Rost's (2011) cognitive processing categories overlap. That is, Field's (2013) lower-level processes which are activated to understand the literal meaning of a text correspond to the neurological and linguistic processing of Rost's (2011) listening processing model. The higher-level processes which according to Field (2013) are activated to understand a text's discourse and implied meaning are in line with Rost's (2011) semantic and pragmatic processing. In addition, although assigned to different categories, the cognitive listening processes in Rost's (2011) and Field's (2013) models appear to be congruent with the cognitive listening processes in Vandergrift and Goh's (2012) cognitive model of L2 listening comprehension, which in turn is adapted from Anderson's three-stage model of cognitive processing for language comprehension (conceptual processing, parsing, and utilization).

The major difference between the three listening models is the acknowledgement of the role of metacognition in listening processing. The two models built on L1 processing and L2 expert listening – Rost's (2011) and Field's (2013) – do not appear to explicitly acknowledge the role of metacognition in listening comprehension processing. Field (2013), in particular, does not explicitly indicate what strategies are activated to assist comprehension processing when gaps in knowledge and comprehension occur.

Field (2013) includes in his higher-level process category a self-monitoring process (comparing a new piece of information with what has gone before to ensure its consistency, and revising judgements on the accuracy of the new items). His framework however does not cover the other metacognitive strategies which Vandergrift and Goh (2012) found to be used by successful L2 listeners, such as anticipating words to be heard or predicting what one will hear based on information brought to consciousness and relevant contextual information, and verifying their prediction. This might be because although Field (2013) sees listeners' use of strategies as important in second language development, at the same time he claims that 'strategies do not form part of expert listening' (p.108). The use of strategies, especially compensation strategies, according to Field (2013), indicates a limitation in language use rather than language ability. Field's claim, however, contradicts Bachman and Palmer (2010), who suggest that strategies or strategic competence is part of language ability. It also deviates from other second language researchers' views (O'Malley et al., 1989; Rubin, 1981; Vandergrift & Goh, 2012), who suggest that learners' use of strategies is one important component of language ability which contributes to success in language communication. To further clarify this point, the following section reviews the role of strategies in L2 comprehension.

2.4.6 Role of cognitive and metacognitive strategies in L2 comprehension processing

A considerable body of research in second language acquisition has indicated that strategy use – both cognitive and metacognitive strategies – plays an important role in L2 comprehension processing (Goh, 2002; Graham, Santos, & Vanderplank, 2008; O'Malley et al., 1989; Rubin, 1981; Vandergrift & Goh, 2012). This is partly because learners have more limited L2 linguistic knowledge as well as contextual and cultural knowledge, which are crucial for comprehension to occur (Færch & Kasper, 1986). Cognitive strategies, such as inferencing and elaboration, are thereby essential to bridge gaps in the knowledge that may occur and increase text comprehension. However, some learners might have developed false beliefs about language learning that negatively affect listening comprehension processing (Færch & Kasper, 1986). For instance, they may think that in order to have a complete understanding of a text, they have to decode and understand every linguistic element in the input. This is not likely to be necessary or possible in listening situations which need rapid and online processing. To successfully understand a text, learners may thus need metacognitive strategies to manage their listening behaviours in order to catch up on what they are listening to.

Previous research on listening comprehension (Goh, 2002; Graham et al., 2008; Ren, 2013; Rubin, 1981) has listed a number of strategies which may enhance L2 listening performance. These include cognitive strategies of two kinds: inductive inferencing and deductive inferencing. Inductive referencing is a learner's guess at meaning based on some hunches from a wide range of textual information. Deductive inferencing is when a listener looks for general rules based on knowledge of their own or another language(s) or based on generalizations from many inductive observations.

However, good language learners, as noted by Rubin (1981), must modify their rules for both inductive and deductive reasoning on a continuous basis to successfully understand the texts they attend to. In addition, they monitor their mental processes and ensure that their processes and strategies are effective (Goh, 2002; Rubin, 1981).

In an investigation of the comprehension processes activated by L2 learners, O'Malley et al. (1989) found some strategies that are beneficial to L2 comprehension processing. These include 1) inferencing, i.e. guessing textual meaning on the basis of several sources, 2) rehearsal, i.e., repeating the names of objects or items that have been heard or practising a longer language sequence, 3) organizing, i.e. grouping information to be retained in ways that will enhance comprehension and retention, and 4) elaboration, i.e. relating new information to information previously stored in memory. In some cases, however, elaborations were shown to interfere with rather than assist comprehension. If a text reminded students of something they knew well, they sometimes got so involved in recalling prior knowledge that their attention wandered from the listening task. Thus, elaboration sometimes had a negative effect if the listeners did not carefully monitor their attention. O'Malley et al. also found that self-monitoring, or being aware of their inattentiveness and consciously redirecting their attention back to the task, had to be activated in order to listen successfully.

In addition to managing comprehension processes and enhancing text comprehension, strategies employed during comprehension tasks, as suggested by past research, show the listening proficiency of listeners. Vandergrift (2003), for example, showed that more skilled and less skilled listeners differ in their use of strategies to complete listening tasks. 36 junior high school students learning French as a second language performed a listening comprehension test. The audio listening was stopped three times and a think-aloud was conducted at those points so that the participants could explain what they were thinking or paying attention to while listening. The results indicated that overall, cognitive strategies (inferencing, elaboration, imagery, translation, repetition, and summarization) were used most frequently and more so than metacognitive strategies (advanced organization, directed attention, selective attention, and self-management, monitoring strategies, and problem identification strategies). However, more skilled listeners appeared to use more metacognitive strategies than less skilled learners. It is thus suggested in this study that more skilled learners have more control of the listening process than less skilled ones.

To sum up, since language processing in L2 is generally not as automated as L1 processing, cognitive and metacognitive strategies are needed to solve comprehension problems and regulate cognitive processes (Færch & Kasper, 1986; Vandergrift & Goh, 2012). The use of strategies, as discussed by Vandergrift (2003), has been found to differentiate between high and low proficiency learners. As a result, in addition to the component of cognitive processes, the use of strategies, as recommended by researchers such as Bachman and Palmer (1996), Phakiti (2008), and Zhang, Goh, and Kunnan (2014) should be counted as part of the construct underlying language ability. Vandergrift and Goh (2012), in particular, have recommended to explicitly include it in L2 listening comprehension models. Following this line of thought, in this study the use of strategies (cognitive and metacognitive strategies) is considered as part of listening comprehension abilities. The next section describes the listening comprehension model that forms the framework of the study, which has been derived from the existing models reviewed above.

2.4.7 The model of listening comprehension processing employed in this study

L2 listening comprehension, as indicated in the listening models described earlier, appears to be restricted by two factors: the level of knowledge listeners possess and the level of expertise or automaticity in processing. Listeners' knowledge, according to Rost (2011) and Field (2013), comprises both linguistic and non-linguistic knowledge. Linguistic or language-related knowledge is a domain of information in an individual's memory used to create and interpret discourse in language use. It includes knowledge of phonology/graphology, lexis, and syntax. These types of knowledge are employed mainly during linguistic processing. They enable listeners to encode speech into linguistic units, detect phonetic features and recognize words in connected speech, and interpret the incoming text. Semantic and pragmatic knowledge is generally activated at a high level of processing, i.e. meaning and discourse construction (Field, 2013). It enables listeners to interpret textual discourse by relating utterances or sentences to each other, to the speaker's intentions, and to characteristics of the language use setting (Bachman & Palmer, 1996).

Another type of knowledge that affects L2 listening is the cultural and world knowledge that listeners bring to listening situations (Field, 2013). Knowledge of this kind has been found to be shaped by listeners' cultural background and experience. It is activated mainly at a high-level of text processing and is especially crucial when listeners have to make inferences or elaborations on the message being delivered (Field, 2013).

Field (2013) mentions that effective L2 listening depends not only on listeners' knowledge but also the degree to which they can process that knowledge automatically. As indicated in the listening comprehension models (see sections 2.4.1-2.4.4), listening ability integrates a number of psycholinguistic abilities. Rost (2011) divides these abilities into four levels: neurological, linguistic, semantic, and pragmatic processing. Field (2013) later put this processing into two main categories according to the levels of cognitive development: lower-level processes or linguistic processing consisting of decoding, word search, and syntactic parsing, and higher-level processes, comprising meaning and discourse constructing. However, due to limitations in knowledge (including both L2 linguistic and non-linguistic knowledge), L2 users employ some strategies to manage their comprehension processing and bridge gaps in their comprehension. Thus, as suggested in previous studies (O'Malley et al., 1989; Rubin, 1981; Vandergrift & Goh, 2012), the use of strategies should be included in the construct underlying listening abilities.

Taking into account what the literature suggests, listening comprehension processes in this study are described in relation to three components, namely cognitive processes, cognitive strategies, and metacognitive strategies (see Figure 2.3). These will form the basis for the analyses of participants' listening processing behaviours during listening-to-summarize task performance.

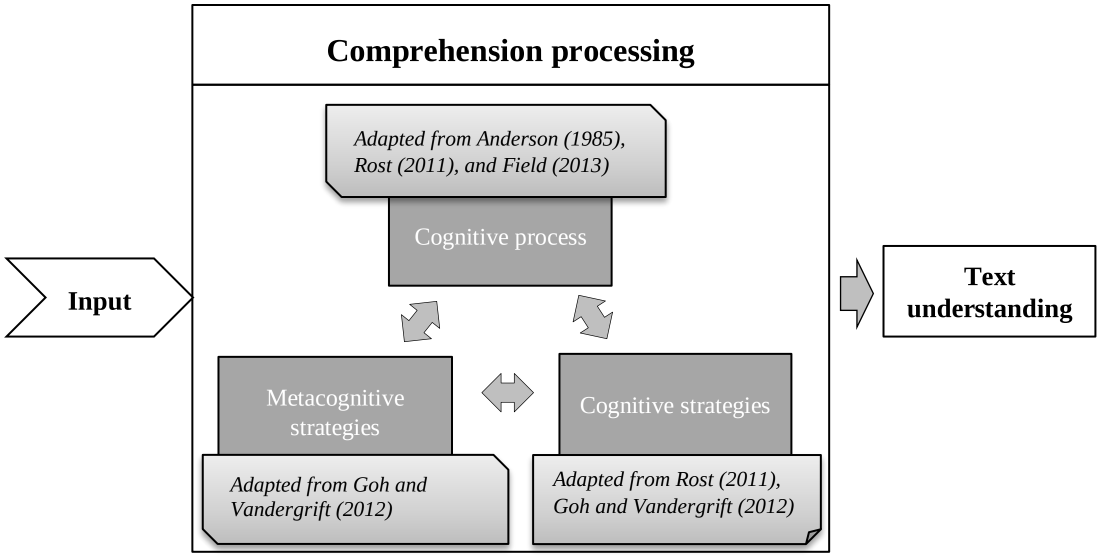


Figure 2.3: Components of listening comprehension processing behaviors

Cognitive processes

In line with Anderson (1985), cognitive processes in this study are considered as a category of mental actions that contribute directly to text comprehension. Incorporating Anderson's (1985), Rost's (2011), and Field's (2013) models, the cognitive processes are sub-divided into six processing types consisting of 1) acoustic-phonetic decoding, 2) word recognition, 3) parsing, 4) semantic processing at the local level, 5) semantic processing at the global level, and 6) pragmatic processing. Acoustic-phonetic decoding occurs when a listener accesses acoustic sounds, registers the sounds, and converts the sounds into the representations of the language phonological system. At this stage of processing, phonemes or phonological forms which are the basic units of words are identified. Word recognition is the process by which the listener segments continuous speech to identify words or series of words (phrases) in a speech stream. Word recognition, as the review suggests, occurs in the form of either bottom-up or top-down processing or the integration of both. Parsing is mapping recognized words onto the syntactic or semantic structures of the language or segmenting chucks of information. The result of a parsing process is propositions, generally consisting of one predicate and one argument (an agent, an object, or a verb modifier) (Anderson, 1985; Kintsch & van Dijk, 1978). Semantic processing takes place when the listener combines the textual information and relates it to their world knowledge to understand the text's meaning. Semantic processing, according to Kintsch and van Dijk (1978), occurs at two different levels, i.e. the local and the global levels. Semantic processing at the local level is creating connections between individual propositions. At the global level of semantic processing or discourse level, the listener links connected propositions to the theme of the text. This is to understand the discourse meaning of the text. However, as the true meaning of a text is often not explicitly stated, listeners have to use their pragmatic knowledge to determine the speaker's intentions. That is, they need to elaborate on the linguistic information and use their social and cultural knowledge about the context of communication. This is referred to as pragmatic processing. Following Field (2013), these six types of cognitive processes are divided into two levels of processing; lower-level and higher-level processes. At the lower-level of processing are acoustic-phonetic processing, word recognition, and parsing, and the higher-level of processing entails semantic processing at both the local and the global levels, and pragmatic processing.

Cognitive strategies

Strategies are different from processes in that strategies are used with some degree of consciousness whereas processes are more automatic (Vandergrift & Goh, 2012). Classifying according to their functions, cognitive strategies refer to the strategies used while listening is going on to solve listening comprehension problems. Based on previous research, e.g., Goh (2002), O'Malley et al., (1989), and Vandergrift (2003), and Vandergrift and Goh (2012), cognitive strategies in this study are described as 1) inferencing, or the use of linguistic information gained in listening to fill in missing information and guess the meaning of unfamiliar words, 2) elaboration, or using background knowledge or topical knowledge to make the text meaningful, 3) prediction, or anticipating listening content, 4) translation, or changing words, phrases or sentences into L1 before interpretation, 5) fixation, or stopping to think or focus attention on understanding a small part of a text, and 6) reconstruction, or using key words to recreate meaning.

Metacognitive strategies

Following Goh and Vandergrift (2012), metacognitive strategies, in this study, concern the strategic competence that provides a management function in language use. These strategies are used to manage and oversee the listening process, including the use of cognitive strategies. Relying on previous research on L2 listening comprehension (Goh, 2002; O'Malley et al., 1989; Vandergrift, 2003; Vandergrift & Goh, 2012), metacognitive strategies involve in general the processes of 1) planning or preparing and analyzing the requirements of a listening task, 2) paying attention selectively to what a listener expects to hear, 3) re-directing attention when it is away from the incoming text, 4) monitoring comprehension or activating the appropriate listening processes, solving comprehension problems, and verifying predictions when they are not accurate according to the text being listened to, 5) real-time assessment of input, and 6) evaluating the listening outcome. Under metacognitive management, mental or cognitive processes and strategies are expected to proceed more efficiently.

Figure 2.4 summarizes the model for listening processing – its components and sub-processes and strategies – used to analyze listening processing behaviours in this study. The model comprises both cognitive and strategic processing. Cognitive processing comprises six cognitive processes activated by listeners to understand listening input. Strategic processing involves the cognitive strategies used to solve comprehension problems or fill gaps in listening comprehension and the metacognitive strategies used to manage and oversee the listening process. This model will form the framework for the analyses of test-takers' processes and strategies to comprehend listening input in listening-to-summarize tasks, in order to shed light on the listening construct underlying these tasks.

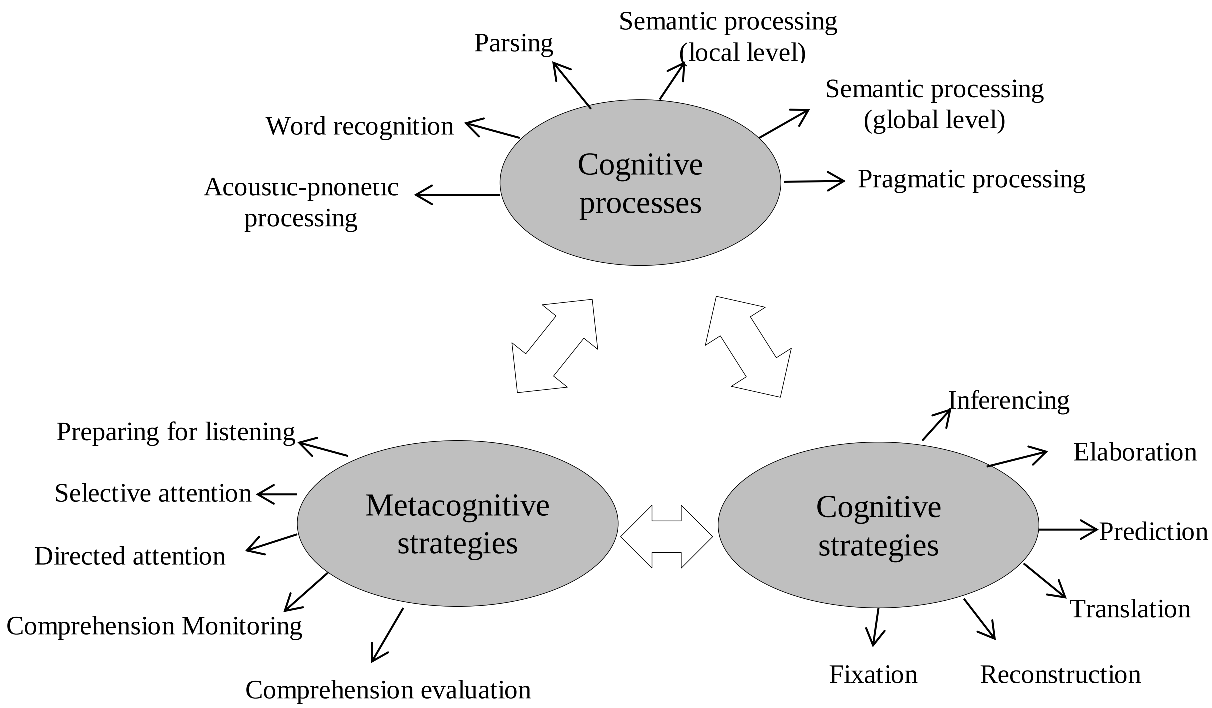


Figure 2.4: Summary of processes and strategies and their sub-components

It is relevant to note, however, that according to Messick (1989; 1995) and Weir (2005), a test construct is not explained just by the cognitive and linguistic abilities within individuals but by the interaction of these abilities in the target language use domain. In practice, there are factors that could affect task performance and score interpretations, such as task authenticity, test fairness, and difficulty. This study therefore has been extended to explore how test-takers perceive these factors and the extent to which their perceived listening task difficulty relates to their performance. The next section reviews the literature related to this.

2.5 Test-takers' perceptions of tasks and of listening task difficulty

Weir (2005), similarly to Messick (1989), emphasizes that it is important that the interaction between test-takers and the context of performance is taken into account in attempting to describe test constructs. In his own words, Weir (2005) shows:

> A test should always be constructed on an explicit specification, which addresses both the cognitive and linguistic abilities involved in activities in the language use domain of interest, as well as the context in which these abilities are performed (theory-based validity and context validity). In our view, construct validity is a function of the interaction of these two aspects of validity and is not just a matter of ability within the individual in isolation (p. 14).

Language performance, according to Douglas (2000), is context-dependent and thus language performance in different contexts requires different abilities. Weir (2005), therefore, suggests that test tasks should be relevant to the target domain of language use as this will indicate the extent to which the test results can be generalized beyond test situations. One source of data which can be useful in this matter and which Weir (2005) suggests in his validation framework is test-takers. Test-takers and their physical, psychological, and experiential characteristics, as Weir (2005) contends, directly affect the way they perform test tasks and thus, as stressed by Weir (2005) and O'Sullivan (2012), test-takers should be at the heart of test design and development to ensure test validity.

Apart from exploring test-takers' listening processing, this study therefore explores test-takers' perceptions of tasks and task difficulty, aiming in particular to describe test-takers' views on task authenticity, fairness, and listening task difficulty and the extent to which the perceived listening difficulty relates to listening performance. It is hoped this evidence will explain some behaviours in listening performance and the extent to which listening-to-summarize tasks are useful for language assessment from a test-takers' perspective. The following section first presents the benefits of investigating test-takers' perceptions in language assessments. Next, the task characteristics that are studied through test-takers' perceptions in this study are outlined.

2.5.1 Use of test-takers' perceptions in language testing

Rost (2011) describes, in reference to listening, that perceptions are the cognitive, cultural and emotional aspects that influence the way a person listens, i.e. senses the world, categorizes and codifies experiences. According to Rost (2011), listeners' perceptions, which are shaped by personal background and experiences, will principally affect the way they participate in listening (i.e. attending to the input, processing text, recalling information, and reporting their understanding of listening events).

In language testing, perceptions have been used to indicate test validity. Traditionally, they are used to indicate face validity or the extent to which a test appears to be an appropriate test and is perceived to be testing what it is said to be testing (Alderson et al., 1995). Face validity, as Alderson et al. (1995) explain, essentially involves the intuitive judgment of people who are not necessarily experts. Perceptions of test-takers on tests in particular are thought to potentially affect the way test-takers respond to and perform the tests. Namely, test-takers are more likely to perform to the best of their ability when they consider tests to be valid; in contrast, they may not take tests seriously when they do not look like proper tests to them (Alderson et al., 1995). These researchers thus recommend future test validation to include test-takers' perspectives.

In the recent conceptualizations of test validity, perceptions are linked to the appropriateness of tasks and the relevance of test tasks to tasks used in real-life situations. In other words, this is to demonstrate the degree to which test tasks, as perceived by test-takers, represent tasks they encounter outside test situations. If test-takers perceive test tasks as being relevant to their real-world tasks, the tasks, as considered by Bachman (2005, p. 26), have 'test appeal' or are perceived to correspond closely to tasks in authentic situations. Weir (2005), in addition, explains that although it is unlikely that tests can cater for all test-takers' individual differences, it is important to be informed about their perceptions of test tasks they have experienced to understand whether they think the tasks are relevant to their real-world experience. This is to put test-takers at their ease, as far as possible, in test situations and to make inferences on their test scores as accurately as possible.

Additionally, test-takers' perceptions have also been relied on to investigate task difficulty, particularly to point out the properties of test tasks which influence performance (Bachman, 2005; Messick, 1995; Weir, 2005; Elder et al., 2002). This data, as Norris, Brown, Hudson and Bonk (2002) found, helps testers to select appropriate test tasks and ensure that the tasks are representative in terms of their difficulty. Test-task difficulty has been investigated in a number of previous studies, e.g., Elder et al. (2002), Iwashita et al. (2001), and Brindley and Slatyer (2002). In most cases, task difficulty was manipulated through the task characteristics such as speech rate, text type, input source, and item format, and it was investigated whether these factors were significantly related to task performance. Nevertheless, results failed to confirm relationships between the manipulations of difficulty and task performance. Brindley and Slatyer (2002), for example, found that some items that the developers had anticipated to be difficult were found to be easy by some test-takers. In line with Messick (1989), these researchers explained that this was because task performance may not be a matter of the tasks themselves but an interaction between the nature of task input, the nature of the assessment tasks, and the individual test-takers.

Considering that task difficulty is not only the result of difficulty manipulated in task input, but of the interaction between tasks, test-takers, and test conditions, studies such as Robinson (2001), Tavakoli (2009), and Révész and Brunfaut (2013) collected test-takers' perceptions/attitudes in order to indicate task difficulty. These studies particularly took the view that the learners/test-takers who are directly affected by the tasks come to test situations with some personal backgrounds which may be difficult for teachers/testers to predict. Hence, learners/test-takers may possess some knowledge that assists or limits task performance and makes items estimated to be difficult by the teachers/testers easy for the learners/test-takers or vice versa.

In Révész and Brunfaut's (2013) study, for example, the researchers compared task difficulty as assessed by test-takers in a perception questionaire to task difficulty as estimated by the Rasch model[4]. 68 participants with different L1 backgrounds performed 18 versions of a listening task. After each of the tasks, they completed a perception questionnaire. The results showed that the participants' perceptions of task and task difficulty correlated strongly to actual task difficulty as assessed by the Rasch model. A

[4] A statistical analysis of task difficulty that combines raw data on individual items and test-takers' abilities to estimate task difficulty

strong correlation between these variables, found in this study, to some extent suggests that test-takers have the potential to accurately estimate task difficulty.

To conclude, this study agrees with previous research that test-takers who have completed tasks should be one of the key data sources for test validiation. They can particularly indicate whether test tasks are relevant to those they encounter in their actual lives. Also, they have been shown to have the potential to indicate task difficulty. Therefore, test-takers' perceptions are a second focus of the analyses in this study. Presented next are the aspects of the tasks that will be investigated through test-takers' perceptions.

2.5.2 The investigation of test-takers' perceptions

The investigation of test-takers' perceptions of tasks in this study relates to three characteristics of integrated-test tasks. They are task authenticity, fairness, and listening difficulty. The first two aspects (authenticity and fairness), as presented in Chapter 1, underlie the use of integrated-test tasks for assessing purposes. Therefore, it seems necessary to explore how test-takers perceive these in listening-to-summarize tasks. In addition, the investigation of perceptions of listening difficulty allows the researcher to look into the cognitive demands of listening input and the extent to which the perceptions of listening difficulty relate to task performance. Each of these characteristics (task authenticity, fairness, and listening difficulty) will be detailed in the following paragraphs.

Task authenticity

One important benefit of integrated-test tasks, as pointed out in previous research such as Cumming et al. (2004), is task authenticity, or the degree of representativeness of the test tasks to tasks in the target domain (Bachman & Palmer, 2010). Task authenticity, as Bachman and Palmer (2010) describe, is important for score interpretations. Authenticity of test tasks to some extent indicates the generalizability of test scores to real-world use of the target language. The interpretations of what test-takers can do in the target situation, as Weir (2005) explains, are more accurate when test scores are obtained from test tasks that represent characteristics of real-world tasks than when they are not. Task authenticity, therefore, as emphasized by testing researchers such as Bachman and Palmer (2010) and Weir (2005) should be shown in test validation.

Test fairness

In addition to task authenticity, the perceptions of fairness – another important benefit of integrated tasks – is investigated. In addition to being authentic, test tasks, as Weir (2005) states, have to be acceptable to test-takers in terms of their fairness or the perceived potential of the tasks to fairly assess abilities. Different test-takers could have different perceptions of test fairness and their perceptions could potentially affect their task performance. Test-takers are likely to perform to the best of their abilities when they consider tests to be fair ways to measure their abilities (Bachman &Palmer, 2011). Therefore, test fairness is focused on in this study.

Task difficulty

Another important characteristic of tasks that should be investigated to reflect task appropriateness is task difficulty. As discussed in section 2.5.2, perceptions of task difficulty are useful to identify task difficulty since task difficulty is not just a result of characteristics manipulated in the task input itself but of the interaction between tasks, test-takers, and the context of performance. This study thus investigates task difficulty, as experienced by test-takers by looking into their perceptions of listening tasks. To achieve this aim, Skehan's (1996, 1998) cognitive complexity framework was used for two reasons. First, it is a comprehensive framework which is widely used to collect data on task difficulty (Frost et al., 2011). Second, text characteristics such as lexical complexity, textual density, and textual organization, as described in this framework, are relevant to those found in previous research to be associated with listening difficulty (e.g., Bloomfield et al., 2011; Gilakjani & Ahmadi, 2011; Révész & Brunfaut, 2013). Following Skehan (1996, 1998), task difficulty is described in relation to three components, namely code complexity, cognitive complexity, and communicative stress.

Code complexity refers to the complexity of the linguistic code itself (the traditional areas of syntactic and lexical difficulty and range, as well as redundancy in and density of the language). Language is simply seen as less-to-more complex in fairly traditional ways. The more complex vocabulary and syntax the task contains, the more difficult the task becomes (Skehan, 1998). Previous studies on listening such as Révész and Brunfaut (2013) and Brunfaut and Révész (2015) have been found to support this view as they show that listening was difficult when listening texts contained unclear pronunciation, difficult lexis, complex grammatical structure, and/or unclear organization.

Cognitive complexity is associated with the content of what is said and comprises the areas of familiarity and processing (Skehan, 1998). Cognitive familiarity is the learners' familiarity with a topic, discourse type, and task types. Learners who are cognitively familiar with tasks are believed to have knowledge which can be retrieved and mobilized for task performance. Cognitive processing, in contrast, is concerned with the extent to which learners have to actively think through the task content. This deals particularly with processing load, which is caused by the organization of information, the amount of online computation, the clarity and sufficiency of information and the information type, such as concrete-abstract or static-dynamic. Listening texts are more difficult to understand when they contain implied meanings and more abstract ideas, and lack discourse markers (Bloomfield et al., 2011).

Communicative stress involves a group of factors unrelated to code or cognitive complexities, but which are thought to restrict communication. It includes such factors as time pressure (how quickly the task has to be done), speed of presentation, number of participants, the length of texts used, the type of response, and the opportunity to control interaction. Taylor and Geranpayeh (2011) indicate that time constraints or time pressure could be a factor affecting listening performance since listening is undertaken in real-time, so time pressure is already built into any listening test task in a way that tends to be less true for reading or writing tasks.

In sum, with the view that test-takers who have experienced performing test tasks and are directly affected by the interpretations of test scores and use are a valuable source of data for test validation, their perspectives on the task used will thus be analysed. Specifically, their perceptions of task authenticity, test fairness, and listening task difficulty will form the focus. Their perceptions of task authenticity and fairness are expected to reveal the extent to which they agree with the rationale underlying the use of integrated-testing in language assessment. Their perceptions of task difficulty, in addition, can provide information on the level of task difficulty which is not easy for testers to estimate and on the extent to which this perceived listening difficulty relates to their task performance. This is hoped to further inform the validity of this task type in language assessment.

2.6 Summary

This literature review has indicated that a major problem in the use of integrated-test tasks to assess L2 performance relates to a lack of clarity on the abilities assessed by these tasks. Particularly in the case of tasks that include listening input, e.g., listening-to-summarize tasks, it remains unclear what listening abilities are involved and measured, and this leads to difficulties in interpretations and use of test scores. Integrated listening tasks are in fact increasingly used in large-scale academic tests such as the TOEFL or PTE Academic. However, very little research has been devoted to their construct, especially where listening is concerned. Therefore, this study set out to describe the listening construct underlying this task type. Informed by Messick (1989) and Weir (2005), the study analyses two sources of data: test-takers' listening processing behaviours and their perceptions of tasks and listening task difficulty. Based on the literature review, listening processing behaviours are described in terms of cognitive processes (the general category of mental behaviours contributing directly to text comprehension), cognitive strategies (specific actions consciously activated to solve comprehension problems), and metacognitive strategies (specific actions consciously used to manage and monitor listening processes). The investigation of perceptions of tasks and listening difficulty concerns three task characteristics, namely, task authenticity, fairness, and listening difficulty. These empirical data are hoped to throw light on the construct of listening-to-summarize tasks, as far as listening is concerned.

The next chapter (3) translates the research gaps described in the present chapter (2) into research questions and presents the methodology employed to answer the research questions.

Chapter 3 Methodology

3.1 Introduction

The purposes of this study are to uncover the cognitive processes and strategies listeners perform in comprehending listening input materials in listening-to-summarize tasks and participants' perceptions of these tasks in three respects, namely 1) task authenticity, 2) fairness and 3) task difficulty (in terms of code complexity, cognitive complexity, and communication stress). The investigation of test-takers' cognitive processes and strategies aims to describe the listening construct underlying the tasks, namely the comprehension part of task performance which has received the least research attention compared to other language skills involved, i.e. speaking and writing. The investigation of test-takers' perceptions, in addition, is to find out how test-takers, who are directly affected by these tasks, perceived them with respect to authenticity, fairness, and difficulty, and whether their perceptions of listening task difficulty affect their task performance. These two research aims are translated into the research questions presented in Section 3.2. Section 3.3 provides a profile of the participants and Section 3.4 presents the research instruments and data collection methods (a background questionnaire, verbal protocols, and perception questionnaires). Pilot studies were carried out to evaluate the suitability of the data collection techniques and will be described in Section 3.5. The data collection procedures of the main study are outlined in Section 3.6, and the final methodological Section (3.7) details the data analyses.

3.2 Research questions

The research aims and gaps indicated earlier can be translated into the following overall research question:

"What listening abilities are assessed by EAP listening-to-summarize tasks?"

In addition, a number of subordinate questions have been formulated:

1. What listening cognitive processes and strategies do ESL test-takers engage in while performing (adapted) PTE Academic listening-to-summarize tasks?
   a. Are there any differences in the processes and strategies when compared between tasks with different language modalities, namely listening-to-speak and listening-to-write, and between performance levels?
2. What are ESL test-takers' perceptions of (adapted) PTE Academic listening-to-summarize tasks and listening task difficulty?
   a. Is the tasks' listening difficulty as perceived by the test-takers related to listening performance?

3.3 Participants

In total, 72 people participated in the study. These were all Thai-L1 speakers who were pursuing university studies at various institutions in the UK. They were invited to take part for four reasons. First, similar to the PTE Academic[5] target population; they were a group of ESL learners who needed English for their academic courses. Secondly, at the

---

[5] The research tasks used in this study were adapted from PTE Academic.

time of data collection (November, 2013 – March, 2014), they were conducting undergraduate or postgraduate studies, representing the study levels for which the PTE Academic has been designed and is typically used as part of language proficiency entry screening. However, as most Thai students in the UK were postgrads, the majority of the participants taking part in this study were postgraduate students. Thirdly, it was hoped their experience in using English for academic purposes would assist them in reflecting on whether the task features and the content and linguistic demands imposed by the tasks represent language use in non-testing situations. The final reason is related to their first language. The research methodology literature strongly recommends allowing participants to use their first language (or at least the choice to do so), in particular when collecting verbal report data, in order to avoid incomplete or vague data due to participants' lack of very high proficiency in the target language (Dörnyei, 2007; Gass & Mackey, 2000). Since Thai is the researcher's first language, English second language speakers with Thai as their first language were invited to take part, so that verbal protocols to look into task processing behaviours could in principle be conducted in the participants' first language and so that the protocols would not require translation for analyses.

In order to collect data to inform the two main research questions, the participants were divided into two groups, A and B. Group A, which was to investigate test-takers' cognitive processes and strategies (RQs 1 and 1a) consisted of 12 participants. Group B, which was to explore test-takers' perceptions (RQs 2 and 2a) contained 60 participants. All participants were given a participant information sheet and informed consent was acquired from each of them.

Group A's participants

The 12 participants in Group A were asked to perform four listening-to-summarize tasks and complete a perception questionnaire and stimulated recall after each task. Purposive sampling was used to engage these participants. That is, the researcher selected the participants based on their English language ability and their availability for data collection. As this data collection aimed to describe the cognitive processes and strategies used by participants with different performance scores, participants with different levels of English language ability had to be included. Therefore, six participants with the minimum entry language score when applying for their study program (6.0 or 6.5 in the IELTS overall band score) and six with higher scores (7.0-8.5) were invited to take part. It should be noted that the IELTS scores (see Table 3.2) are only a very rough indication of their English proficiency. These scores were self-reported by the participants and were obtained prior to entering their study program. For many this was for no more than one year prior to the data collection. Thus, it should be kept in mind that these scores are likely to underrepresent the participants' ability, but it was felt better to use this rough indication than no indicator at all to try and obtain a mixture of English abilities in the participant group. Unfortunately, administering a full proficiency test to all Thai students at Lancaster University (and other UK universities) as part of the participant recruitment process was not feasible.

These participants were recruited by sending out an invitation letter (Appendix 1) and a short survey on IELTS scores to Thai students at Lancaster University in the academic year 2013-2014. The first 12 students who responded with IELTS scores that met the research criteria and who were willing to participate effectively took part in this qualitative data collection, which took about 2.5-3 hours. Half of the participants were male, half were female (see Table 3.1). Their age ranged between 23-40 years old, with a mean of 29. Six were doing a Masters degree and six were PhD students. They were studying in two different faculties, i.e. Arts and Social Sciences, and the Management School, with a variety of subject areas: Applied Linguistics, Linguistics, Law, Design for Sustainability, Marketing, E-Business and Innovations, Management, International Business, HR consulting, and Entrepreneurship, Innovation, and Practice. Their English proficiency, as indicated by IELTS scores (see Table 3.2), ranged between 6.0 -8.5 on the overall band, and 5.5-7.5 in listening, 5.5-8.5 in reading, 5.5-8.0 in writing, and 5.5-8.5 in speaking.

Group B's participants

To study test-takers' perceptions, in a way that also allows for statistical analyses, another 60 participants were invited to participate; specifically to complete listening-to-summarise tasks and perception questionnaires. To reach a sufficient number, these were recruited from four different UK universities: 47 from Lancaster University, five from the University of Edinburgh, four from the University of Birmingham, and four from the University of Bedfordshire. They were one foundation year student, one undergraduate, 45 Masters, and 13 PhD students.

Somewhat over half of the participants (61.7%) were female, whereas 38.3% were male (see Table 3.1). Their age ranged between 20-40 years old, with a mean of 27. At the time of data collection, they were studying in six faculties with various subject areas:

1) Arts and Social Sciences (Arts Design, Applied Linguistics, TEFL, TESOL, Media and Cultural Studies, Gender and Women Studies, and Law), 2) Management School (Marketing, E-Business and Innovation, Management, International Business, HR consulting, Entrepreneurship, Innovation and Practice, and Accounting and Financial Management), 3) Engineering (Chemical Engineering), 4) Science/Science and Technology (Materials, Health, Safety and Environment), 5) Tourism (Project Management and Tourism management), and 6) School of Education. Over half of the participants (63.3%) were studying in a Management School. Their English proficiency (see Table 3.2), as indicated by their IELTS overall scores, was 4.5-8.0[6]. Their scores on individual skills ranged between 4.5-8.0 in listening, 5.0-8.5 in reading, 4.5-7.5 in writing, and 5.5-8.5 in speaking.

Participants' background		**Group A (n=12)**	**Group B (n= 60)**
Age (ranged in years)		23-40 ($\bar{x}$= 29, S.D.=5.56)	20-40 ($\bar{x}$= 27, S.D.=4.29)
Gender	Male	6 (50%)	23 (38.33%)
	Female	6 (50%)	37 (61.67%)
Level of study	Foundation Year	0	1 (1.67%)
	Undergraduate	0	1 (1.67%)
	Master	6 (41.6%)	45 (75.00%)
	PhD	6 (58.4%)	13 (21.67%)
Faculty	Arts and Social Sciences	5 (41.67%)	15 (25.00%)
	Management School	7 (58.33%)	38 (63.33%)
	School of Education	0	2 (3.33%)
	Science and Technology	0	2 (3.33%)
	Engineering	0	1 (1.67%)
	Tourism	0	2 (3.33%)

Table 3.1: Data on participants' backgrounds

[6] The participants whose IELTS overall score was lower than 6.5 (about 25.0%) had taken pre-sessional English courses before entering their study program and before the data collection took place.

Participants' English proficiency		IELTS score bands/ No. of participants 4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	Missing
Group A (n=12)												
Overall band score		0	0	0	3	3	2	2	1	1	0	0
Individual band	Listening	0	0	2	1	2	1	2	0	0	0	2
	Reading	0	0	1	3	3	1	1	1	2	0	0
	Writing	0	0	2	2	4	1	2	1	0	0	0
	Speaking	0	0	1	3	3	2	0	1	2	0	0
Group B (n= 60)												
Overall band score		1	1	4	13	22	10	5	2	0	0	0
Individual band	Listening	2	1	10	10	9	9	6	9	0	0	2
	Reading	0	2	9	11	14	8	6	4	3	0	1
	Writing	1	3	11	19	14	7	2	0	0	0	1
	Speaking	0	0	6	19	13	8	5	4	2	0	1

Table 3.2: Participants' English proficiency demonstrated by IELTS scores

3.4 Data collection methods and instruments

This section describes the research methods and instruments used for data collection, comprising a background questionnaire (3.4.1), listening-to-summarize tasks (3.4.2), verbal protocols (3.4.3), and perception questionnaires (3.4.4). Ethical approval for the research design was gained from Lancaster University's Ethics Committee.

3.4.1 A background questionnaire

To gain a more precise profile of the participants, their bio-data were collected by means of a background questionnaire. The questionnaire (see Appendix 2) consisted of 10 items eliciting information on the participants' age, first language, gender, educational background, overseas experience (specifically in an English speaking context) and their English language proficiency.

3.4.2 Listening-to-summarize tasks

As explained before, the tasks under investigation in this study are integrated-test tasks with a listening component; more specifically, listening-to-speak and listening-to-write tasks. In practice, a total of eight listening-to-summarize tasks were used, adapted from the PTE Academic (http://pearsonpte.com/). The PTE Academic is an English proficiency test aimed at non-native English speakers who need to demonstrate their academic English ability for university admission or professional purposes. The basis for the tasks used in this study were two PTE Academic *Re-tell Lecture* items (originally listening-to-speak tasks), and two *Summarize Spoken Text* items (originally listening-to-write tasks). The tasks required the participants to first listen to an academic lecture and after the listening to provide a summary in either oral or written form. As will be explained in more detail below, adaptations were made to the original tasks to make the two task types more similar to allow for comparison (see below points a) and b)), to have a counterbalanced design, and to be able to look into any potential impact of the productive requirements of the task (speaking vs. writing; see point c)).

a) Instructions changed from '*re-telling*' to '*summarizing*' a lecture

To investigate the effect of different modalities (speaking and writing) on listening comprehension processing, two different types of tasks were used: *Re-tell Lecture* items involving listening and speaking skills, and *Summarize Spoken Text* items involving listening and writing skills. However, as these original tasks seemed to require different forms of responses, i.e. 're-telling' versus 'summarizing', a pilot study was conducted to find out whether and how the different task instructions affect performance behaviours (see 3.5). The pilot study results showed that when different wordings were used in the instructions, the participants interpreted the requirement of the tasks differently. In the tasks with the instruction to summarize the listening text, the pilot participant understood that he had to select only the message(s) relevant to the text's key point. When the tasks instructed test-takers to retell the listening text, the participant thought that he had to rephrase all the information he heard without thinking whether it was relevant to the main point. Because of these results, it was thus necessary to opt for one of the two instructions to allow for comparisons between listening-to-speak and listening-to-write tasks. 'Summarize' was used in this study because, compared to 're-tell', it requires more skills necessary for listening comprehension, including selecting and organizing information (see 3.5).

b) Visual input added

The original version of the *Summarize Spoken Text* items did not provide an image for the test-taker, whereas the *Re-tell Lecture* items did. Therefore, in the adapted tasks, an image was added to the *Summarize Spoken Text* items so that their task input was more comparable to the *Re-tell Lecture* items, and any potential impact of different productive skills (speaking and writing) on listening comprehension processing could be investigated.

In the original *Re-tell Lecture* items, the picture provided indicated the topic of the listening passage and illustrated some information that listeners had to link to the aural text in order to fully understand the idea underlying it. Thus, the images added to the *Summarize Spoken Text* items were related to the listening input and reflected the topic of the listening passage and some key ideas which listeners had to interpret on the basis of what they listened to.

c) Parallel tasks developed

After the adaptation of the original tasks was completed, another four parallel tasks were developed. More specifically, for the two PTE Academic tasks that originally required an oral summary, two alternative tasks were developed, which required a written summary, and vice versa. Thus, in essence, the listening input remained exactly the same, but the way in which the participants had to provide the summary was altered (oral vs. written). This was to investigate if there were any differences in listening comprehension processing behaviours when different modalities (speaking and writing) are required in task production.

As can be seen in Table 3.3, this resulted in a total of eight listening-to-summarize tasks, developed from what were originally four listening input topics, i.e. *Corruption*, *Hans Krebs*, *Talent*, and *Vitamin D*.

Tasks		Listening Topic	Length of listening (seconds)	Skills involved	Prep. Time	Response Time
Adapted tasks	1	Corruption*	1:29	Listening-Speaking	10 (Seconds)	40 (Seconds)
	2	Hans Krebs*	1:10	Listening-Speaking	10 (Seconds)	40 (Seconds)
	3	Talent**	1:25	Listening-Writing	---------10 (Minutes)---------	
	4	Vitamin D**	1:01	Listening-Writing	---------10 (Minutes)---------	
Parallel tasks developed	5	Corruption	1:29	Listening-Writing	---------10 (Minutes)---------	
	6	Hans Krebs	1:10	Listening-Writing	---------10 (Minutes)---------	
	7	Talent	1:25	Listening-Speaking	10 (Seconds)	40 (Seconds)
	8	Vitamin D	1:01	Listening-Speaking	10 (Seconds)	40 (Seconds)

*Originally *Re-tell Lecture* items; **Originally *Summarize Spoken Text* items

Table 3.3: Listening-to-summarize tasks

The listening input length ranged from 1:00-1:29 seconds. Based on each listening input two forms of response were required, i.e. oral and written summaries. The tasks that required an oral summary are referred to as listening-to-speak tasks in this thesis, whereas those requiring a written summary are listening-to-write tasks. The tasks were presented to the participants in a Power Point Presentation (PPT), which was timed and set to play automatically when the participants clicked on the start button. Following are screenshots showing how the tasks were delivered to the participants.

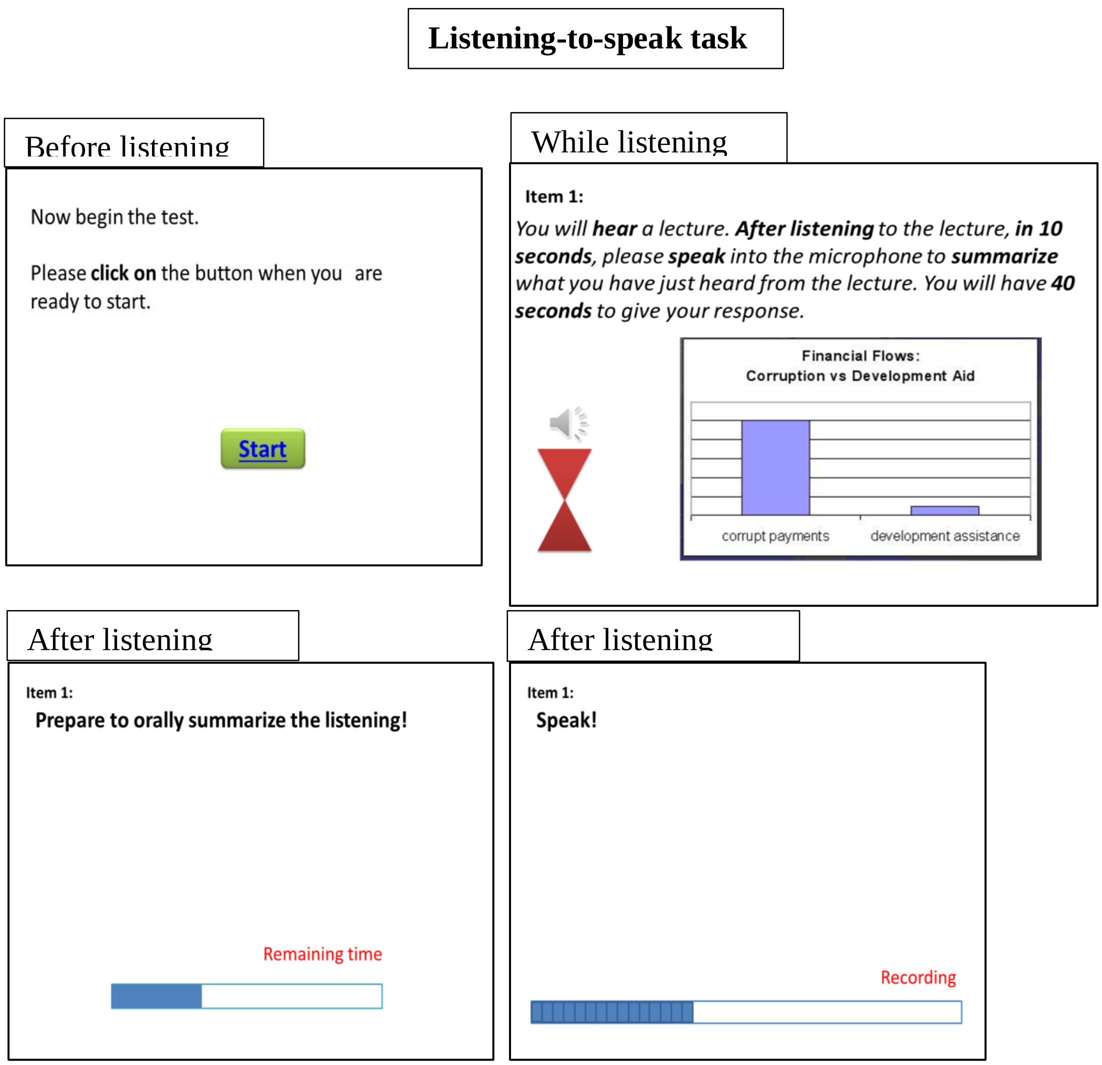


Figure 3.1: Sample of listening-to-speak task slides

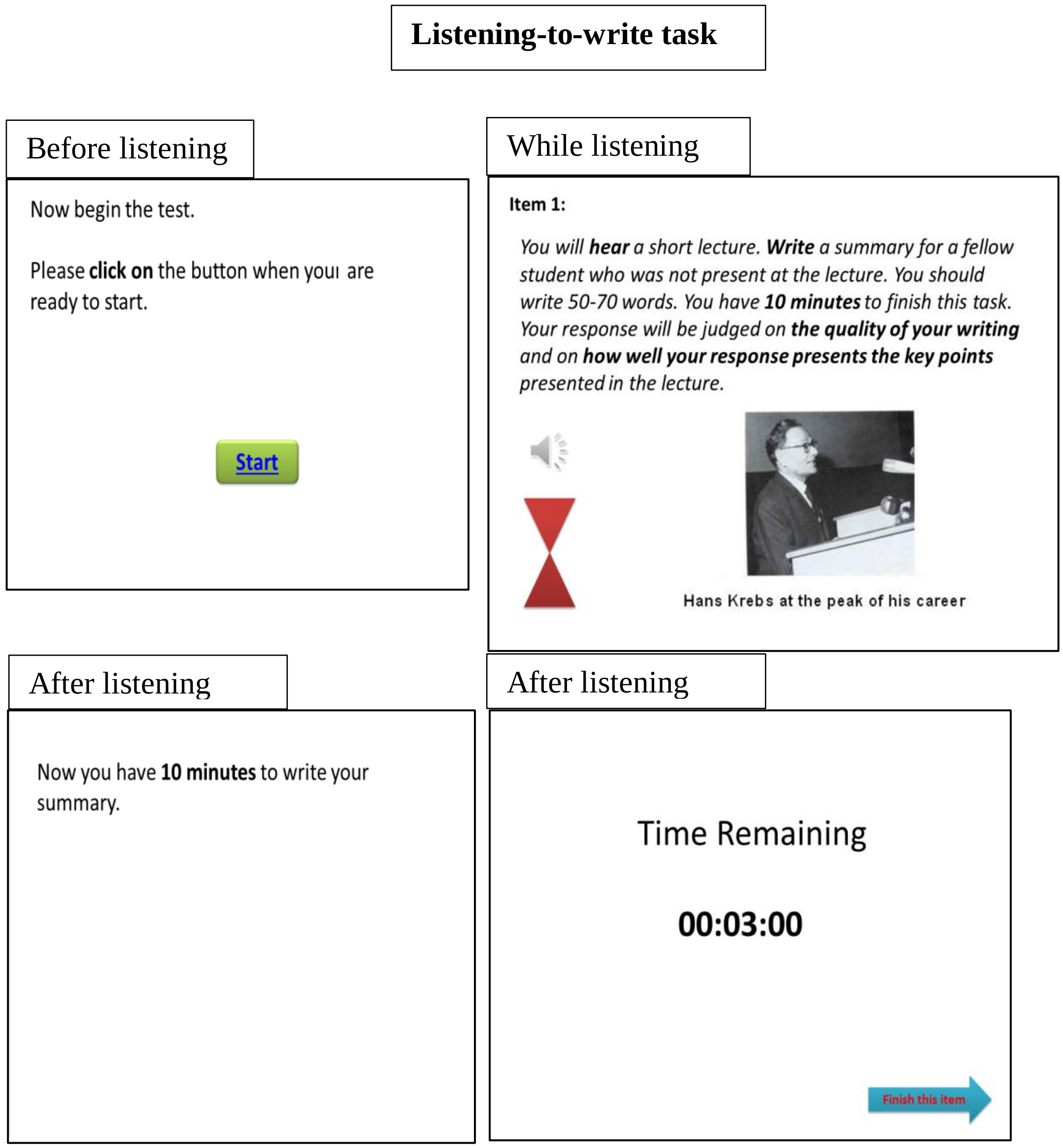


Figure 3.2: Sample of listening-to-write task slides

Following the PTE Academic guidelines, the listening-to-speak tasks allowed participants 10 seconds after listening to prepare and 40 seconds to orally summarize the listening. On the other hand, in the listening-to-write tasks, the participants had 10 minutes to prepare and write a 50-70 word summary of the listening passage. As is the case for the original PTE Academic items, under both conditions, the participants were allowed to take notes whilst listening and a timer was displayed on the screen to show the remaining time for task completion. To familiarize the participants with the item types and to help reduce their test anxiety, two sample items were first given to each participant, which were disregarded for analyses.

3.4.3 Verbal protocols

To investigate the cognitive processes and strategies the participants used to complete the tasks (RQs 1 and 1a), this study employed verbal protocols, i.e. stimulated recalls. Verbal protocols are “oral records of thoughts, provided by subjects when thinking aloud during or immediately after completing a task” (Kasper, 1998, p. 358).

In information-processing theory, verbal reports are considered a useful data source for investigating cognitive processes and strategies (Ericsson & Simon, 1993). Ericsson and Simon (1993) recommended two forms of verbal reports which they claim can closely reflect individuals’ cognitive processes and strategies and experiences. Divided according to the period when information is accessed, these are concurrent verbal reports and retrospective reports. Concurrent verbal reports, or talk-aloud and think-aloud protocols, are the direct verbalization of information heeded during actual cognitive processing or when task performance is going on (Ericsson & Simon, 1993). Verbalizing thought processes at this period is not a description or explanation of what one is doing, but verbalization of what one is paying attention to while generating answers to the task. Retrospective protocols, on the other hand, are the reports of cognitive processes which have (just) finished (Ericsson & Simon, 1993). Immediately after task completion, some retrieval cues are thought to remain in short-term memory (STM), allowing individuals to recall their thought processes with supposedly high accuracy and completeness. This is especially the case, as Ericsson and Simon suggest, when tasks take less than 10 seconds to complete. However, the longer the period between task processing and retrospective reporting, the more difficult and incomplete the recall. Therefore, stimuli that can help individuals to recall their thought processes are recommended to be included (Ericsson & Simon, 1993), such as a video recording of task performances, notes taken by the participants, or task output (Gass & Mackey, 2000).

Another form of retrospective protocol is retrospective verbal reports or interviews, requiring participants to explain or describe the processes they performed during task performance after they have completed the task. Unlike stimulated-recall, no prompts are given to stimulate the participants' verbalization. Instead, specific questions are asked to guide the reporting, for example, a) 'What were you focusing on when you responded to this situation?' and b) 'What made you reply in this manner?'(Ren, 2013).

A number of empirical studies (Goh, 2002; O'Malley et al., 1989; Ren, 2013; Vandergrift, 2003) have demonstrated that verbal protocols can provide useful information on cognitive processing in the context of language processing and also on the strategies users employ to complete language tasks. Verbal report data, in addition, can provide evidence of sources of knowledge applied to complete tasks (Goh, 2002). Table 3.4 lists studies that have used one or more of the techniques to investigate learners' and test-takers' cognitive processes in language task performance. Focusing on the studies investigating listening processes, it was found that verbal reports were one of the main data sources; see for instance, in Yi'an (1998) and Vandergrift (1997; 2003). However, each of these three techniques has its own strengths and limitations, as summarized in Table 3.4.

In essence, the key advantage, as found across the three techniques (think-aloud protocols, stimulated recall, and retrospective verbal report) and listed in Table 3.4, is that they seem to reveal processes and strategies used to complete the tasks which are not otherwise directly observable by the researcher(s). Studies that have used the techniques (see Vandergrift, 1997; 2003; Field, 2012) have found that verbal protocol data evidenced strategies and processes activated successfully and unsuccessfully by the participants. In addition, they revealed the knowledge sources participants use to complete tasks. However, the distinctive benefit of think-aloud protocols (versus the other two techniques), as pointed out by O'Malley et al. (1989), is that these can tap into cognitive processes and strategies that may otherwise be lost in retrospective techniques such as stimulated recalls or retrospective report/interview techniques due to the time lapse.

Study	Data aimed to collect	Tasks used	Language skill(s) focused	Strengths found in previous studies	Limitations speculated in previous studies
1) Think-aloud protocols					
O'Malley (1989)	Cognitive processes and strategies L2 learners used	Lecture listening	Listening	• Well demonstrated strategies used by participants (Plakans, 2009). • Revealed how writers with different proficiency levels differed in their decision making and use of strategies (O'Malley et al., 1989; Plakans, 2009; Vandergrift, 1997, 2003). • Cognitiv processing which was lost in retrospection was described (O'Malley et al., 1989). • Tapped into naturally occurring behaviours (Storey, 1997).	• Limited to the participant's ability to articulate information (O'Malley et al., 1989). • The act of verbalizing distracted from task performance (writing task) (O'Malley et al., 1989; Plakans, 2009; Storey, 1997). • Participants may provide only a limited range of strategies of which they were consciously aware while performing the task (O'Malley et al., 1989; Storey, 1997). • When not reported in the participant's L1, thinking–aloud protocol data was somewhat incomplete (Weigle et al., 2013).
Storey (1997)	Cognitive processes test-takers engaged in	A multiple-choice discourse-cloze test	Reading		
Vandergrift (1997, 2003)	Strategies used by L2 French learners	Multiple-choice comprehension questions	Listening		
Goh (2002)	L2 listeners' comprehension strategies	Listening texts in various topics	Listening		
Plakans (2009)	Reading strategies used by non-native English writers	An integrated reading-writing task	Reading		
Weigle et al. (2013)	Cognitive processes involved in short-answer questions	A reading-writing task	Reading		

Study	Data aimed to collect	Tasks used	Language skill(s) focused	Strengths found in previous studies	Limitations speculated in previous studies
2) Stimulated recalls					
Yi'an (1998)	Test-takers' test taking processes	Multiple-choice questions	Listening	• Provided insightful data on cognitive processes (Field, 2012; Yi'an, 1998). • Pointed out to different strategies used by successful test-takers (Swain et al., 2009). • Pointed out to knowledge used in comprehension process and interaction of different listening tactics used (Goh, 2002).	• Could possibly not capture some processes activated by the participants (Field, 2012) (Barkaoui et al., 2013). • Might not tap into some processes automatically performed by the participants (Swain et al., 2009). • Not able to provide information on whether the strategies were effective at the production stage (Barkaoui et al., 2013).
Vandergrift (1997; 2003)	Strategies used by L2 French learners	Multiple-choice comprehension questions	Listening		
Goh (2002)	L2 listeners' comprehension strategies	Listening texts in various topics	Listening		
Swain et al. (2009)	Processes and knowledge test-takers used to complete integrated and independent speaking tasks	• Independentspeaking • Integratedreading-listening-speaking • Integrated listening-speaking	Speaking		
Field (2012)	Cognitive processes used in listening performance	• Listening comprehension • Lecture listening	Listening		
Barkaoui et al. (2013)	Strategic behaviours test-takers reported using when responding to integrated and independent speaking tasks	• Independent speaking • Integrated speaking tasks	Reading Listening Speaking		

Study	Data aimed to collect	Tasks used	Language skill(s) focused	Strengths found in previous studies	Limitations speculated in previous studies
3) Retrospective interview/report					
Vandergrift (1997; 2003)	Strategies used by L2 French learners	Multiple-choice comprehension questions	Listening	• Allowed the researcher to follow up the point unclear (Field, 2012). • Permitted analysis of learners' attention to information and factors influencing their production (Ren, 2013). • Pointed out successful and unsuccessful use of strategies (Graham et al., 2008).	• Some processes may not be reported by participants (Ren, 2013).
Goh (2002)	L2 listeners' comprehension strategies	Listening texts in various topics	Listening		
Graham et al. (2008)	Development of strategy use over 6 months listening	Multiple-choice listening task	Listening		
Field (2012)	Cognitive processes used in listening performance	• Multiple-choice listening tasks • Lecture listening	Listening		
Ren (2013)	Cognitive processes of advanced L2 learners during their study abroad programme	• Listening to audio-recordings of different topics	Listening		

Table 3.4: Techniques used to collect verbal protocol data in previous studies investigating cognitive processes and strategies

Despite their advantages, concerns have been raised over the use of verbal reports as research data (Barkaoui, Brooks, Swain, & Lapkin, 2013; Crutcher, 1994; Ericsson & Simon, 1993). As verbal reports rely mainly on individuals' access to information in their own memory and their ability to verbalize such information, concerns over veridicality or the extent to which the verbal information accurately reflects thought processes have been expressed (Ericsson & Simon, 1993). For example, verbal data can be accurate only when individuals are truthful or reporting exactly what they were thinking while performing the tasks and not reporting things that they think the researcher(s) want to hear. Also, because some processes occur automatically and are unlikely to be available to the participant's consciousness, some processes may have been activated but not articulated by participants, resulting in incomplete data. This risk has been discussed in the literature on both listening and reading comprehension (e.g., O'Malley et al., 1989; Swain et al., 2009), and some evidence was found in the case of studies employing think-aloud protocols (see O'Malley & Chamot, 1990), stimulated recalls (see Barkaoui et al., 2013; Swain et al., 2009), and retrospective interviews (see Ren, 2013). Furthermore, the ability of individuals to verbalize information and in particular their proficiency in the language used for verbalization were also found to affect the accuracy and completeness of information (O'Malley et al., 1989; Weigle et al., 2013). It is thus recommended to use the participants' first language in data collection (if at all possible).

A second issue is the risk of reactivity, especially with respect to think-aloud protocols. It has been found that the act of thinking-aloud during task performance may affect or alter the performance. O'Malley et al. (1989) specifically found that the requirement to perform tasks and at the same time report one's processing appeared to alter participants' task processing. The act of verbalizing during writing task performance in particular, as noted in O'Malley et al. (1989), Plakans (2009), and Storey (1997), disturbed the writing task performance. As a result, the scores obtained might or might not represent participants' actual ability.

A further issue concerning verbal protocols relates to the reliability and validity of data coding. The verbal data obtained from these techniques are not a direct report of the thoughts or cognitive processes that the participants performed in completing tasks but the information they attended to during the tasks (Ericsson & Simon, 1993; Kasper, 1998). Hence, in order to arrive at the processes involved, the researchers have to make inferences on the basis of the data provided by the participants. When protocol data are not carefully analyzed (and, for example, no coding reliability is established), the validity and reliability of the results are highly questionable (Kasper, 1998).

In language testing research, verbal reporting data have been used to investigate the construct underlying tasks. Despite the potential weaknesses, the methods have been found useful for analyzing processes and strategies used in task performance (Barkaoui et al., 2013; Plakans, 2009; Storey, 1997; Weigle et al., 2013). As mentioned by Barkaoui et al. (2013), however, protocols are not able to provide information on whether the strategies are effective during the task performance. They are unlikely to explain whether the processes and strategies used contribute either to success or failure in task output. Therefore, to reveal the effectiveness of each process and strategy, Barkaoui et al. (2013) have advised also analysing test-takers' task output.

In summary, the literature review has indicated both the benefits and disadvantages of verbal protocols as a data source for investigating cognitive processes and strategies activated to complete tasks. On the positive side, the method has been shown to lead to useful insights into cognitive processes and strategies, which are otherwise difficult to observe. In language testing studies in particular, previous research suggests that verbal reports provide useful information on test-takers' cognitive processes and strategies, which is crucial to describe the construct underlying the tasks in question effectively. However, the accuracy and completeness of verbal data can be affected by a number of factors, including the automaticity of cognitive processes, individuals' abilities to articulate the information heeded, the language used for verbalization, the time period between processing and verbalization, and issues with the data coding process. Nevertheless, careful research design and procedures for verbal protocol data collection and analyses can help minimize the potential risks of the methods (Green, 1998). Therefore, since the aim of this study was to describe the listening construct by investigating test-takers' cognitive processing of listening (RQ 1), it was decided to use a verbal report method to collect data. At the same time, great care was taken to try and avoid some of the pitfalls of this method.

In practice, a stimulated-recall technique was used for three main reasons. First, the task responses in this study were scored for the participants' language ability. Stimulated recalls, which were conducted after task completion, were considered appropriate to minimize the effect of the data collection technique (if any) on task performance. Second, as it was necessary to collect data after task completion, stimulated recalls were considered important since they provided participants with some stimuli (a video recorded during the task performance and the notes taken during the listening) to stimulate their thought processes. Third, as the study concerns research on listening, using a think-aloud would likely be very disruptive since participants have to talk while trying to listen.

3.4.4 Perception questionnaires

As discussed in Chapter 2, data on test-takers' perceptions of tasks and task difficulty can inform task validity in terms of task authenticity, fairness, and task difficulty. Research Question 2 therefore asked what test-takers' perceptions of these three aspects are. To answer this question, a questionnaire was used for a number of reasons. First, questionnaires, as indicated in previous research (Tavakoli, 2009), benefit an investigation of test-takers' perceptions at least in two ways. One is related to the fact that questionnaires can be designed to tap into the perceptions of different factors which the literature suggests condition task difficulty (Tavakoli, 2009). The other is that questionnaires, specifically with multi-item scales, allow researchers to fully capture various traits underlying an individual factor (Dörnyei, 2007). Secondly, this study aimed to explore the relationship between perceptions of listening task difficulty and task performance to consider whether they are related to each other. Since questionnaires, particularly if they use closed-response items, can provide quantitative data (versus for example qualitative data gathered by interviews), this method seemed usefully able to investigate the relationship between task perceptions and performances. In addition, in order to know how the participants perceived the tasks, especially in terms of task difficulty, it was preferable to collect the data immediately after task completion in order to obtain as accurate as possible data. Furthermore, since some participants (Group A) were required to do stimulated recalls on their task processing behaviors, a quick questionnaire after task completion and before recall was preferable over, for example, more time-consuming interviews to keep the time gap between performance and recall limited.

Two different questionnaires were used. One was designed with reference to the listening-to-speak tasks (see Appendix 3) and the other for the listening-to-write tasks (see Appendix 4), because the productive side (speaking versus writing) implied some differences. That is, two statements were version-specific (Statements 14 and 21) and worded to be accurate to the productive skill involved in task performance (writing vs. speaking). This was done particularly to investigate whether different modalities required after listening make participants' perceived listening difficulty different. Each questionnaire aimed to measure the participants' perceptions in three components, namely 1) perceptions of task authenticity, 2) task fairness, and 3) task difficulty (in three aspects: code complexity, cognitive complexity and communication stress) (see Table 3.5). These three foci had been opted for because, as mentioned in Chapter 2 (section 2.5), the investigation of test-takers' perceptions aimed to provide data for test validation in terms of task authenticity, fairness and listening task difficulty, all of which is crucial to inform the usefulness of the tasks for assessment purposes.

To measure each individual factor, multi-item scales (meaning that more than one item was used to tap into the same aspect) were employed. This was in order to ensure comprehensive coverage of each aspect.

Questionnaire construct	No. of items	Item No.
1) Task authenticity	2	1,2
2) Task fairness	3	3,4,5
3) Task difficulty		
Code complexity	5	6,7,8,9,10
Cognitive complexity (cognitive familiarity)	4	11,12,13,14
Cognitive complexity (cognitive processing demand)	5	15,16,17,18,19
Communication stress	4	20,21,22,23
Total	23	

Table 3.5: The perception questionnaire construct

In total, each questionnaire consisted of 23 Likert-scale statements on which participants had to indicate their level of agreement on a five-point scale (5-strongly disagree, 4-agree, 3-neutral, 2-disagree, and 1-strongly disagree). The two versions of the questionnaire (one for the tasks requiring a written summary and the other for those requiring an oral summary of the listening) mainly contained the same statements. Both versions of the questionnaires were provided to the participants in their first language, i.e. Thai. The questionnaires were administered immediately after a task had been completed and the participants completed the same list of questions for each individual task.

3.5 Pilot studies

Prior to collecting the data, three pilot studies were carried out. Pilot study 1 was to pilot the research task materials. Pilot study 2 was to ensure that stimulated recalls were able to provide data that answer the research questions and for the researcher to gain experience in using this particular method. The perception questionnaires were tried out in Pilot study 3. In what follows, a summary is given of each pilot study and a description of the amendments (if any) made as a result of piloting to the plan for data collection in the main study.

Pilot study 1: the research task materials

In this pilot study, four original tasks from the PTE Academic (see 3.4.2) were given to a Thai student studying towards a Master's degree at Lancaster University. This student was invited to take part because he was considered to represent the study's target population (see 3.3). After completing the tasks, the participant was asked to watch a video recording of his task performance and explain 1) how he had understood the instructions, and 2) how he had approached the tasks and what he was thinking about/ paying attention to during the tasks. This showed that in the *Re-tell Lecture* items, the participant understood that he had to retell or restate as many pieces of information gained from the listening as possible. The organization of the ideas, as understood by the participant, was not as important as it was in the *Summarize Spoken Text* items. Unlike the *Re-tell Lecture* items, in the *Summarize Spoken Text* items, the participant thought that it was very important to organize the information that he understood from the listening. As explained by the participant, he would include in his summary only key and relevant information. Given this difference in interpretation of what was required by each task, it was decided to revise the *Re-tell Lecture* instructions and require a summary (oral) after listening instead of retelling. This was to control for the effect of task instructions on task performance behaviours in order to be able to compare the effect of differences in productive medium (speaking/writing) more directly. The choice in this study was to opt for a summary in both task types (rather than retelling in both) because previous research (Field, 2012; Yu, 2013) showed that summaries require test-takers to engage in the higher-levels of listening processing such as semantic and discourse constructing.

In addition, the pilot participant commented that the visual information given in the *Re-tell lecture* items helped him to understand the points in the spoken text quickly and thought that the use of visuals reflected lectures that he attended at the university. Thus, to keep the two task types similar and enable comparisons, visual information was added to the adapted versions of the *Summarize Spoken Text* items. As the pictures in the *Re-tell lecture* (the *Corruption* and *Hans Krebs* texts) provide the topic and a couple of phrases hinting at key information of the text, similar information was given in the *Summarize Spoken Text* tasks (the *Vitamin D* and *Talent* texts).

Pilot study 2: usefulness of and experience with stimulated recalls

Four Thai students studying towards their degree at Lancaster University participated in this pilot study. One was an undergraduate student in Management BBA, and three of them were postgraduate students (one an MSc in Finance, one an MA in English Language and Literacy Studies, and one a PhD in Economics). Similar to Pilot study 1, these participants were invited to take part because they were considered to represent the study's target population.

These participants were first asked to complete a background questionnaire and then performed four adapted PTE Academic listening-to-summarize tasks (two listening-to-speak and two listening-to-write tasks). After each task, they conducted a stimulated recall, which took about 20 minutes. In the stimulated recall procedure, the participants were presented with a video recording of their task performance and invited to explain what they were thinking about or paying attention to during task completion. In the first round of playing the video, the participants were told to stop the recording wherever they wanted to describe their thought processes during their task performance. In the second round, the researcher stopped the recording at three different time points and invited the participants to explain what they were doing or paying attention to in each period. With some participants, the recording was played three to four times in order to stimulate their verbalization and gain insightful data.

The pilot study data were analyzed using Field's (2013) cognitive processing framework for listening, one of the dominant frameworks used in analyzing L2 listening processes. The analysis revealed that different cognitive processes were activated by the participants to comprehend the listening input, including parsing and meaning construction. Although some of the processes, i.e. word decoding, were not verbalized by some of the participants, inferences could be made that the participants engaged in these processes on the basis of the notes and the summaries they had produced. In fact, this was the case when processes occurred automatically, and the participants were not likely to be aware of their own processing behaviors. In line with previous research, this pilot study suggested that automatic processes performed by participants were not captured by verbal protocols.

Pilot study 2, in addition, pointed to limitations in using Field's (2013) framework to analyze verbal data for this study. Although Field agrees that listeners' use of strategies is an important part of L2 listeners' success, it is not entirely clear how he integrates the use of strategies in his framework. This might be because the framework, as explained by Field, was drawn up on the basis of competent users of the target language (English), who have sufficient knowledge to understand texts and need relatively few strategies to solve their comprehension problems or to manage their listening processes. In this pilot study, however, the analysis showed that the participants, even the successful listeners, used some strategies not integrated in Field's framework, such as prediction, selective attention, and directed attention in their listening comprehension processing. In addition, it was found that some strategies, although they were included in Field's framework, appeared to be too narrow to describe the listeners' cognitive behaviors found in this study. For example, 'inferential' is described in Field (2013) as 'the listeners supplied details that the speaker has not felt it necessary to include' (p.101). In this study, however, it was found that listeners inferred not only the information left unsaid by the speaker but also details that are stated but which they failed to recognize.

Based on this pilot study, some changes were therefore made to the main study. First, in addition to the use of stimulated recall data only, an analysis of participants' handwritten notes made whilst listening and their summary content was included in an attempt to describe the listening comprehension processing more comprehensively. This was particularly because the stimulated recalls might not be able to provide evidence of some processes occurring automatically. It was also because of the concern that by the time the participants had completed the summary, they might not be able to comprehensively think back to the way in which they had processed the listening text even though the video recording was used as a stimulus. Second, given the limitations found in Field's (2013) framework, it was decided to develop a framework for verbal data analysis based on several existing cognitive frameworks for language comprehension, including Anderson (1985), Rost (2011), Field (2013), and Vandergrift and Goh (2012), in order to capture the processing behaviors more fully (see 2.4).

Pilot study 3: perception questionnaires

To gain insights into the perception questionnaires' construct validity, the questionnaires and a description of their intended underlying construct was given to a PhD student working in second language acquisition who had extensive experience with questionnaire design. This student indicated whether she thought the questionnaire items tapped into the construct. A few points of disagreement on word choices were then discussed and changes were made where relevant. Unfortunately, it was not possible to pilot the questionnaire, which was in Thai, on a large number of people and conduct statistical analyses at this stage, because that would have meant using up the participant pool for the main study.

In addition, to ensure that the Thai version of the questionnaire conveyed the same meaning as was aimed for in the original English version (which had been developed on the basis of the literature and discussed with the researcher's supervisor and above-mentioned PhD student), the questionnaire was piloted with a Thai-L1 student doing his Masters degree in English Language and Literature at Lancaster University (UK) who was also highly proficient in English. This student was asked to read the questionnaire in Thai and explain how he understood it, item by item. As a result, some ambiguous items were altered.

In summary, three pilot studies were conducted prior to the main study data collection with the aim of trying out the research task materials, the stimulated recall method, qualitative data analysis, and the perception questionnaires. Based on the findings of these pilot studies, amendments were made to the plan for data collection in the main study. This included changing the instructions of the research tasks to require oral summaries instead of retelling the lecture, the provision of visual information in the adapted tasks, changes to the framework used for data analysis, and changes to the questionnaire items. The next section describes the data collection procedures of the main study.

3.6 Data collection procedures in the main study

Before describing in more detail how the data in the main study were collected, an overview of the data collection methods employed to answer the research questions is provided in Table 3.6.

Research questions		Participants	Data collection
RQ1	What listening processes do ESL test-takers engage in while performing (adapted) PTE Academic listening-to-summarize tasks? a. Are there any differences in the processes and strategies when compared between tasks with different language modalities, namely speaking and writing, and performance levels?	Group A (12 Participants)	• Stimulated recall data • Summary content • Listening notes • Perception questionnaire
RQ2	What are ESL test-takers' perceptions of (adapted) PTE Academic listening-to-summarize tasks and task difficulty? a. What is the relationship between the perceptions of listening task difficulty and task performance?	Group B (60 Participants)	• Perception questionnaire • Task performance scores

Table 3.6: Research questions and an overview of data collection methods

To answer the first research question and its sub-question, qualitative data on processes and strategies performed to complete the tasks were collected from a total of 12 participants (Group A). More precisely, the data comprised verbal report data, gathered by means of stimulated recalls, and language data, taken from notes written down by the participants while listening and summaries produced after listening. To answer the second set of research questions, concerning the perceptions of tasks and listening task difficulty and the relationships between the perceptions of listening task difficulty and test-takers' performance, quantitative data were collected from a total of 60 participants who were not involved in the stimulated recalls. The data comprised responses to perception questionnaires and task performance scores.

Figure 3.3 gives an overall picture of how the tasks were delivered to the participants and how the data were collected. As indicated in section 3.2, the participants were separated into two groups, A and B. In both groups, the procedure began with completing a background questionnaire and then completing four listening-to-summarize tasks on a one-to-one basis. After each task performance, the participants in Group A were asked to complete a perception questionnaire and then a stimulated recall, whereas in Group B only a perception questionnaire was administered.

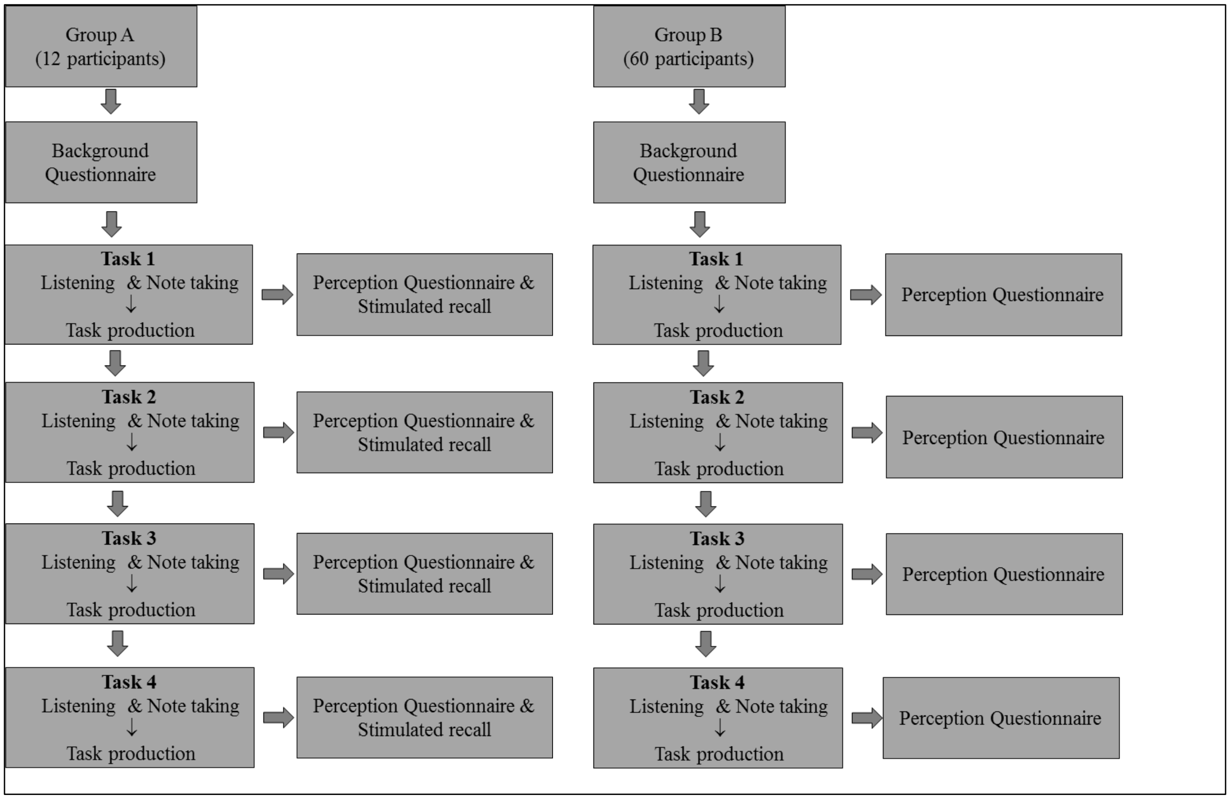


Figure 3.3: An overall picture of the main study's data collection procedures

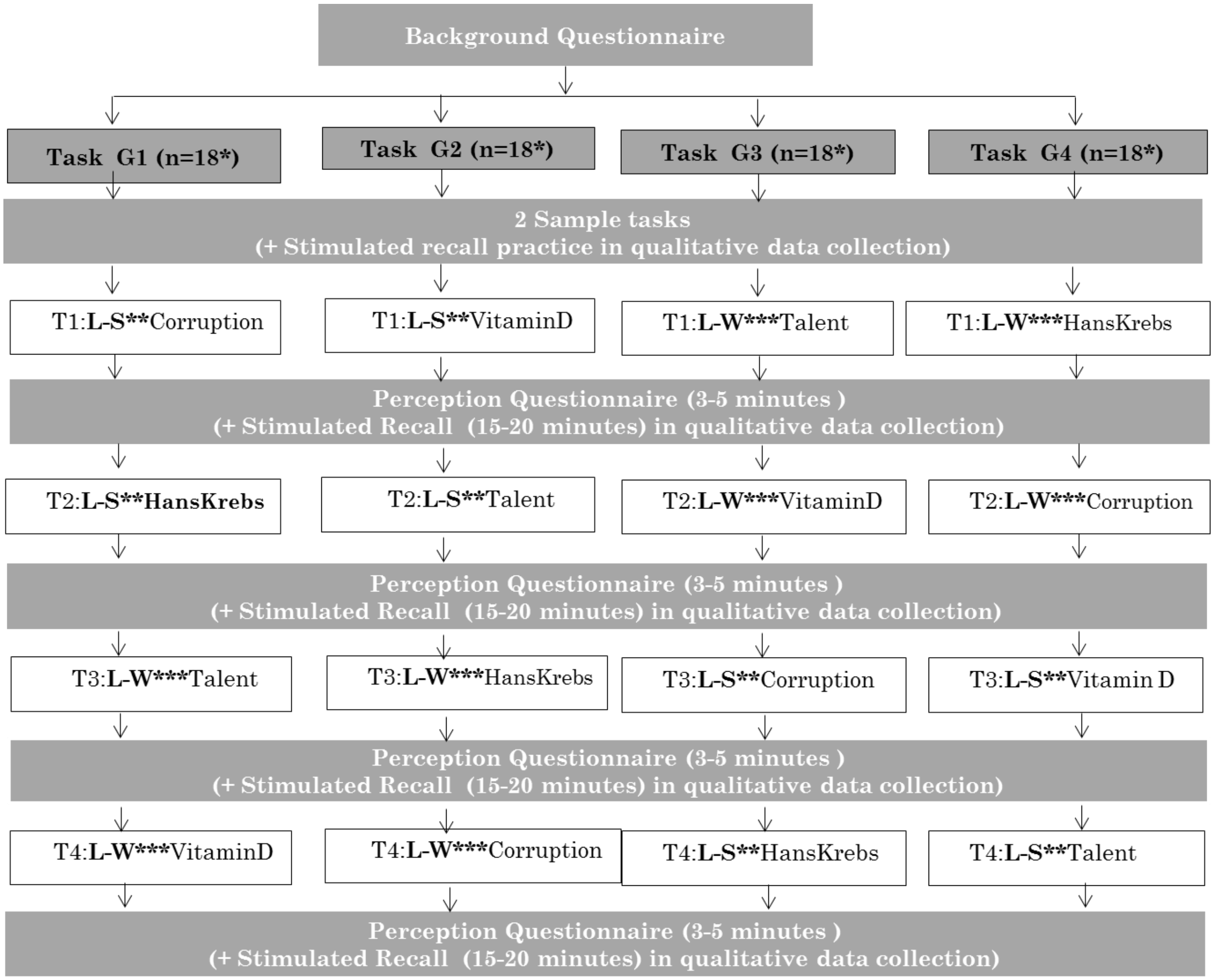
The following explains the data collection step-by-step.

1) Completion of the background questionnaire

To begin with, a background questionnaire was administered to each participant. The questionnaire (see Appendix 2) consisting of 10 items aimed to collect bio-data on the participants' first language, gender, educational background, overseas experience, and their English ability.

2) Completion of the listening-to-summarize tasks

Although indicated earlier that eight tasks were used in the study (see 3.4.2), each participant was asked to perform only four tasks with different listening passages. That is, they performed one of the task groups presented in Figure 3.4, Task Groups 1-4. The grouping of the tasks was based on the listening passages and the skills involved in task performance. To investigate whether different modalities (speaking and writing) affect listening comprehension processing (RQ 1a), it was important that each participant performed both listening-to-speak and listening-to-write tasks. However, to avoid that they heard each listening input more than once so that their experience of producing a summary in one modality would affect a second summary, each task had to be associated with a different listening input. Consequently, in each task group, no listening passage was used more than once. Task Groups 1 and 3, as shown in Figure 3.4, were made up of the same tasks but presented in a different order, and Task Groups 2 and 4 were also made up of the same tasks but presented in a different order. This was in order to minimize the potential effects of task sequencing on task performance. The difference between Task Groups 1+3 versus Task Groups 2+4 was the modality of each task, i.e. listening-to-speak tasks in Task Groups 1+3 were listening-to-write tasks in Task Groups 2+4 (and vice versa).



** 3participants from Group A and 15 participants from Group B*
***L-S = Listening-to-speak tasks & ***L-W = Listening-to-write tasks*

Figure 3.4: Task delivery and data collection

Before completing the four actual tasks, the participants were asked to do two sample tasks (one listening-to-speak and one listening-to-write), in order to familiarize themselves with the item type and to reduce test anxiety. Participants were assigned to a Task Group in such a manner that each Task Group included participants with different levels of language ability (as based on their self-reported IELTS scores).

3) Completion of the perception questionnaire

The perception questionnaire was administered immediately after the participants completed each task. Thus, since the participants performed four different listening-to-summarize tasks, each participant completed the perception questionnaire four times (two times for listening-to-speak and two times for listening-to-write tasks).

4) Participation in stimulated recall (for Group A's Participants)

In Group A, where 12 students participated, a stimulated recall was conducted immediately after the participant completed each task and the perception questionnaire. As each participant was asked to perform four tasks, four stimulated recalls were obtained from each person. The participants were offered a choice of using English and/or Thai for the stimulated recalls. This was to compensate for the impact of language ability on expressing thoughts or opinions. As all the participants used Thai, translations are given in the quotes in the findings chapters (4-6). Prior to the actual data collection, the participants were provided with an opportunity to practice the stimulated recalls on the sample tasks.

For the purpose of stimulated recalls, the 12 participants from whom verbal protocol data were collected were video-recorded whilst completing the tasks. Immediately after they completed each task, they were asked to watch their video-recording and encouraged to describe what they were thinking about or paying attention to while listening to the texts. Similar to the pilot study procedure, the recording was played at least twice in order to obtain insightful data, depending on the participants' abilities to report their own thoughts. In the first round, the participants were in control of pausing and explained whatever they wanted to share. In the second round, the researcher stopped the video recording systematically at the same points for all participants and asked the participants to explain their thought processes. For a couple of participants, the video was played three to four times if they appeared to be quiet during the first and second laps of video playing.

3.7 Data analyses

The data obtained were analysed, following two lines of enquiry.

1. To answer RQs 1 and 1a on the cognitive listening processes and strategies used in task performance, the stimulated recall data, written notes while listening, content summaries, and the questionnaire data obtained from the 12 participants of Group A were analysed.
2. To answer RQs 2 and 2a dealing with perceptions of tasks and listening task difficulty and the relationship between listening task difficulty and task performance, the task performances were scored and the questionnaire data from the 60 participants of Group B were analysed. A correlational analysis was then carried out.

3.7.1 Analysis of stimulated-recall data

The stimulated recalls which the participants conducted in their first language, Thai, were transcribed and analysed to identify processes and strategies activated during listening. To avoid data loss that might occur in translation, stimulated recall data analysis was carried out on the Thai transcriptions. Translations into English were however made to present a selection of the results in this thesis.

The stimulated recall data were analysed qualitatively by the researcher and an external coder. The data were coded using a coding scheme drawn from the literature on listening comprehension processing both in L1 and L2 listening and in learning and testing contexts. The use of a coding scheme has been recommended by Kasper (1998), particularly when researchers need to make inferences as is the case in this study in which participants indirectly verbalized cognitive processes and the researcher has to infer these from participants' reports of heeded information. As Kasper (1989) emphasized, such a coding scheme should be based on a principled and theory-grounded model. In this study, as indicated in section 2.4, the model for analysing task processing behaviours draws on Anderson's (1985), Rost's (2011), Vandergrift and Goh's (2012) and Field's (2013) models. The coding scheme used is described below.

Coding scheme

The coding scheme used in the analysis (see Appendix 5) was composed of three main categories, namely cognitive processes, cognitive strategies, and metacognitive strategies. Each category was sub-divided into different processes/strategies. Under the cognitive processes category fell the subcategories acoustic-phonetic decoding, word recognition, parsing, semantic processing at the local and discourse levels, and pragmatic processing. The cognitive strategy category included fixation, inferencing, elaboration, prediction, translation, and reconstruction. Metacognitive strategies consisted of the subcategories pre-listening preparation, selective attention, directed attention, comprehension monitoring, real-time assessing of input, solving comprehension problems, and comprehension evaluation.

Data coding

Following Gass and Mackey's (2000) suggestions on stimulated recall data analyses, the data were categorized into episodes by using the computer-assisted qualitative data-analysis software NVivo. The data were first segmented into what appeared to be plausible units that corresponded to processes and strategies listed in the coding scheme. Categories of processes/strategies were then assigned to the chunks. For example, the following extract obtained from a 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 listening] didn't say anything about 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) shows that while listening and trying to understand the text, the participant recognized a word (handicraft) in speech. This chunk corresponded to and was categorized as word recognition. In the second chunk (Chunk_2), the participant was establishing links between the words/phrases she had recognized in order to understand the literal meaning of what she had been listening to in *Hans Krebs*. This chunk corresponded to and was classified as semantic processing at the local level. Another example of the segmenting is:

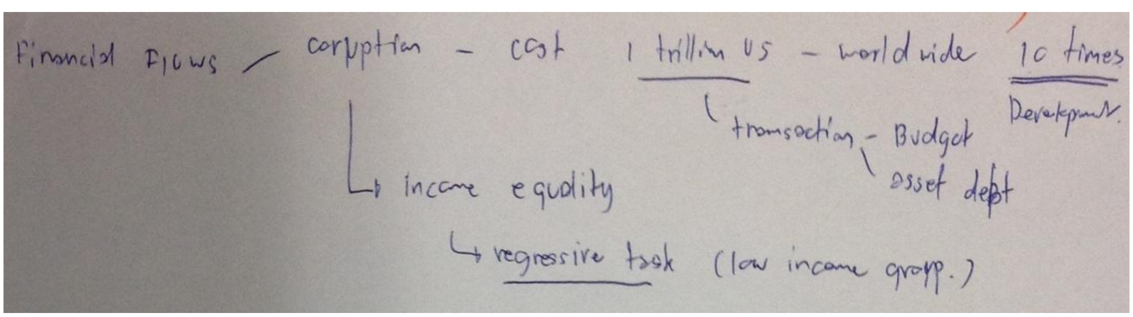
> [Chunk_1] Here I was predicting that the speaker was going to talk about the definitions of talent because he said before, [Chunk_2] 'different ways of defining things restrictive, broad, and meaningless'.

In this piece of data, two chunks were identified. The first chunk (Chunk_1) indicated that the participant used the words he obtained from the beginning of the text to predict the text coming next in the listening. This behaviour corresponded to and was classified as a cognitive strategy, i.e. prediction. The second chunk (Chunk_ 2) shows that during the listening the participants were able to recognize and at the same time group words into a meaning unit of the listening text. This, thus, was categorized as parsing.

3.7.2 Analysis of participants' notes and summary content

To supplement the verbal report data, the participants' notes and the summary content (i.e. of the oral or written summaries produced for the listening-to-summarize tasks) were analysed. Language data in the hand-written notes and summaries were coded using the same coding scheme as used for the stimulated recall analyses. This was in order to investigate processes/strategies that might not be reported in the recall data and/or to confirm those processes/strategies reported. The participants' notes were first analysed and cognitive and strategic behaviours were then inferred from what the participants had written down. The following is an example of a participant's note in the main study and a description of how it was analysed.

Note 3.1



(P12/*Corruption*)

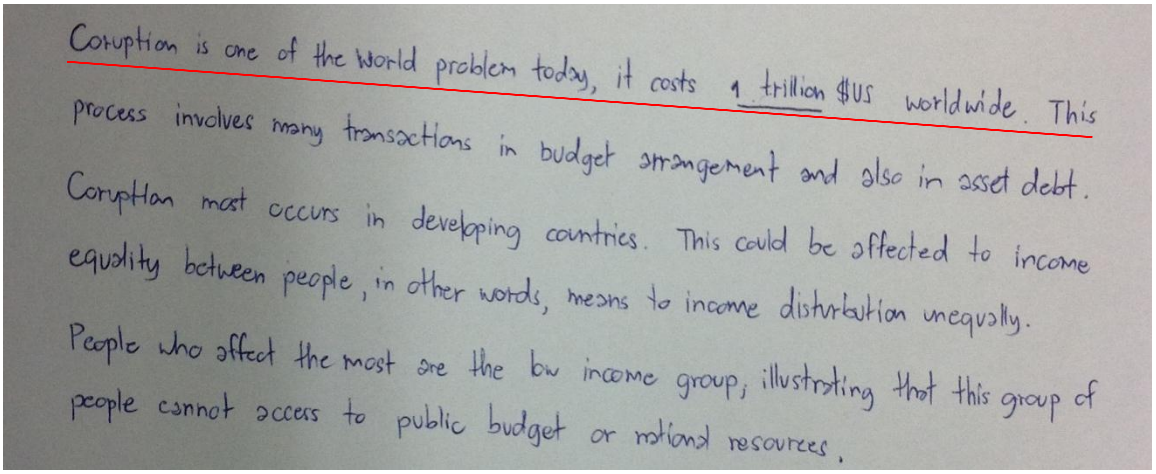
From this note, it was inferred that the participant engaged in at least two processes, namely acoustic-phonetic decoding and word recognition as they had noted a number of words from the listening. The way the words are organized in the note seems to suggest that the participant was trying to relate information while listening, so it was hypothesized from the note that this participant might have also engaged in parsing and semantic mapping or building a mental model of what is heard. These processes were not confirmed until the note was checked against the verbal recall data and the summary produced by the participant after the listening. As in the stimulated recall, this participant explained (see the following excerpt) that he drew a line to link groups of word together.

Quote 3.1

> I listened. I noted down the words I heard, trying to figure out what the story was about. I was trying to relate words together to make a story. I looked at 'corruption' and I draw a line to 'cost'. He said Corruption cost.. how much a year? one trillion dollar I wrote it down….and it happened in 'developing countries and then it affect income inequality' (P12/*Corruption*).

The participant's content summary was then analysed to ensure that he actually engaged in semantic processing and did it successfully. In his summary, the participant wrote:

Written Summary 3.1



Corruption is one of the World problem today, it costs 1 trillion $US worldwide. This process involves many transactions in budget arrangement and also in asset debt. Corruption most occurs in developing countries. This could be affected to income equality between people, in other words, means to income distribution unequally. People who affect the most are the low income group, illustrating that this group of people cannot access to public budget or national resources.

(P12/*Corruption*)

The analysis of the participant's written summary (see the first line) indicates that this participant indeed used the lines from his notes to link pieces of information together as he wrote in the summary that 'Corruption is one of the world problem today, it cost 1 trillion $US worldwide.' This participant's engagement in semantic processing, as previously hypothesised, was thus confirmed. Further analysis of his summary showed that the participant also performed semantic processing at the discourse level, as he could provide the main point of the listening correctly. That is, 'corruption creates income inequality and mainly affects the poor more than the wealthy.' The analyses suggest that this participant engaged in semantic processing at the discourse level and hence the task used was able to tap into this.

Inter-coder reliability

To ensure reliability of the coding process, the stimulated recall data, written notes taken during listening, and listening summary from four participants, i.e. 25% of the data, were double coded by an external coder. As the data were collected in Thai, this second coder was an experienced Thai university lecturer with a Masters degree in Teaching English to Speakers of Other Languages (TESOL) and experience in verbal data analysis. To familiarize the coder with the study materials, the coder was given the coding scheme and one set of sample data, consisting of verbal report data, the participant's notes and listening summary. The researcher discussed the coding scheme and how the data should be technically coded with the external coder, prior to actual data coding.

For the actual data coding, the coder was asked to complete the table for data analysis, drawn up for this study's purposes and adapted from Gass and Mackey (2000) (see Appendix 6). To be specific, after reading through the stimulated recall transcript, the coder was required to put the chunks of data indicating processes/strategies in one column and the category of the processes/strategies they belonged to in the next column. To analyse the participants' notes and the content summary, the coder was asked to identify the types of processes and strategies used on a scanned copy of the notes and summary.

To calculate inter-coder reliability, the coded data from the three different sources (the verbal reporting data, the notes, and the summary) were put together for each participant per task. The two coders (including the researcher) then summarized what processes and strategies each participant engaged in. The evidence on the engagement in cognitive processes were transformed into nominal data by the researcher and two categories were assigned, namely 'use' or 'not use' in order to identify processes that the participants engaged in. The use of cognitive and metacognitive strategies was counted for frequency. A statistical analysis (Cohen's Kappa) was carried out to indicate inter-coder reliability. As presented in Table 3.7, the inter-coder agreement on the overall use of processes and strategies was .772 and the inter-coder agreement on the performance of each aspect of processing behaviours, i.e. cognitive processes, cognitive strategies, and metacognitive strategies, was 0.886, 0.863, and 0.850 respectively. Although overall there was high agreement on the coded data, the two coders discussed those cases where there was disagreement on processes/strategies identified, e.g., how the use of comprehension monitoring was different from real-time assessment of input, until they came to an agreement. The researcher then analysed the rest of the data, i.e. from the remaining eight participants, independently.

Task processing categories		**Level of agreement**	
		Cohen's Kappa	**Sig.**
Process	Cognitive processes	.886	.00
Strategy	Cognitive strategies	.863	.00
	Metacognitive strategies	.850	.00
Overall use		.772	.00

Table 3.7: Analysis of inter-coder reliability

3.7.3 Analysis of test performance

The task responses from the 72 participants (12 from Group A and 60 from Group B) were scored. As each participant had been asked to perform four listening-to-summarize tasks (two listening-to-speak and two listening-to-write tasks), a total of 288 task responses (144 oral summaries and 144 written summaries) were scored. Two experienced PTE Academic raters independently scored all the task responses, using the human rater version of the PTE Academic scoring criteria.[7] The oral summaries were scored on three aspects: content, pronunciation, and fluency. The written summaries were marked on content, grammar, and vocabulary. However, for the purposes of this study, only the content scores were used to evaluate the participants' listening ability as these were taken to form an indicator of the participant's comprehension of listening content.

The content scores obtained from those participating in the qualitative data collection (Group A) were used to evaluate the participants' listening abilities (moderate or highly-able listeners) to conduct further sub-analyses in terms of cognitive processing (see Chapter 7). The content scores from the quantitative data collection (Group B) were used to investigate the relationship between task performance and perceptions (see Chapter 8).

3.7.4 Analysis of the questionnaire data

The responses to the perception questionnaires were statistically analysed using SPSS. As indicated earlier, multi-item scales were used in the questionnaire in order to tap into perceptions of task authenticity, fairness, and listening difficulty. The internal consistency of the questionnaire items was first analysed to assess the homogeneity of the items measuring the same trait. For this purpose, factor analysis was carried out (see Chapter 8 for the results).

To explore the perceptions of tasks and listening task difficulty, descriptive statistics, namely frequencies and percentages, were calculated for all questionnaire

---

[7] There is also an automated rating version.

statements. To investigate the relationship between the perceptions of task difficulty and task performance, a correlational analysis (Pearson correlation) was carried out between Group B's responses on the perception questionnaire items and their content scores on the listening-to-summarize tasks.

3.8 Summary

In this chapter, the research design and the rationale underlying the data collection techniques were explained. The overarching question addressed in this study is what the listening construct underlying listening-to-summarize tasks is. Two major aims were then set: 1) to investigate the listening processes the participants performed in understanding the listening input in the listening-to-summarize tasks, and 2) to explore the participants' perceptions of tasks and listening task difficulty and their relation to task performance. 72 Thai students pursuing undergraduate and postgraduate degrees at universities in the UK took part in this study. They were divided into two groups, A and B. To address the first research aim, stimulated recall data and language data (notes taken during the listening and summaries of the listening) from 12 participants in Group A were collected and analysed qualitatively. To achieve the second research aim, task performances and perception questionnaire data from 60 participants in Group B were collected and analysed statistically. The next chapters report the results of the analyses.

Chapter 4 Cognitive processes

4.1 Introduction

This chapter and the following three chapters report the findings on the participants' cognitive processes and strategies activated to understand the listening input of the listening-to-summarize tasks. This data provides an answer to the first research question: what listening processes do ESL test-takers engage in while performing (adapted) PTE Academic listening-to-summarize tasks? The overall picture of the cognitive and strategic processing behaviours and the number of the participants who performed each of the processes and strategies are described in three chapters: Chapter 4 focusses on cognitive processes, Chapter 5 on cognitive strategies, and Chapter 6 on metacognitive strategies. After that, Chapter 7 compares the processes and strategies described in chapters 4-6 according to task modality and performance level. It is however important for the readers to bear in mind when reading these chapters that the results are presented on the evidence obtained from test-takers. The analysis of these results describes the patterns of evidence presented. To say, for example, that 'only one participant engaged in a process/strategy' should be understood as meaning only one participant presented evidence of engaging in that process/strategy.

As discussed in Chapter 2, listening in L2 is a complex process, involving both cognitive processes and cognitive and metacognitive strategies (Rost, 2011; Vandergrift & Goh, 2012). With regard to cognitive processes, successful listeners, who can understand the main point of what they listen to, engage not only in lower-level processes (acoustic-phonetic processing, word recognition, and parsing) but also higher-level processes (semantic and pragmatic processing) (Field, 2012). Field (2012), and Taylor and Geranpayeh (2011), consequently, have stressed that tasks used to assess listening comprehension ability need to capture both the lower- and the higher-level processes in order to fully represent the construct. This chapter therefore specifically looks into the cognitive processes performed by the participants in comprehending the four listening passages in the listening-to-summarize tasks under investigation, i.e. *Corruption, Talent*, *Vitamin D*, and *Hans Krebs.*

4.2 Cognitive processes

Following the literature reviewed in Chapter 2, cognitive processing in this study is separated into two stages of processing; namely lower-level and higher-level processing. Processing at the lower-level involves linguistic processing in three respects, i.e., acoustic-phonetic processing, word recognition, and parsing (the segmentation of an utterance according to syntactic or semantic cues). The outcome of the processing at this level is meaningful units or chunks of information. At the higher-level of language processing, where comprehension takes place, listeners go beyond the literal meaning of a text to what is more meaningful. This involves semantic processing at the local and the global level and pragmatic processing. At the local level of semantic processing, listeners understand each individual concept of the text either by linking parsed information together or by using surrounding text and/or background information to arrive at it. Semantic processing at the global or discourse level, on the other hand, helps listeners to relate single concepts to the theme of a topic and understand the text meaning as a whole, or, in other words, understand the text's discourse meaning.

The analysis of the stimulated recall transcripts, the participants' notes, and the summary content from the 12 participants in Group A revealed that their cognitive processing in understanding the listening materials falls within these two stages of comprehension processing. Most evidence for this was found in the participants' stimulated recalls. However, as discussed in section 3.5, some processes, especially lower-level processes, might occur automatically, and thus may not have been recognized or reported by the participants. The participants' notes and the summary content were therefore analyzed to supplement the stimulated recall data, hoping that these would reveal information on the more automated, lower-level processes. Specifically, in the presentation of the participants' notes and written summaries, visual annotations (lines and circles) have been added by the researcher to make the points presented clear to the readers.

Figure 4.1 presents the range of processes found and the number of participants performing the processes while listening to each audio file.

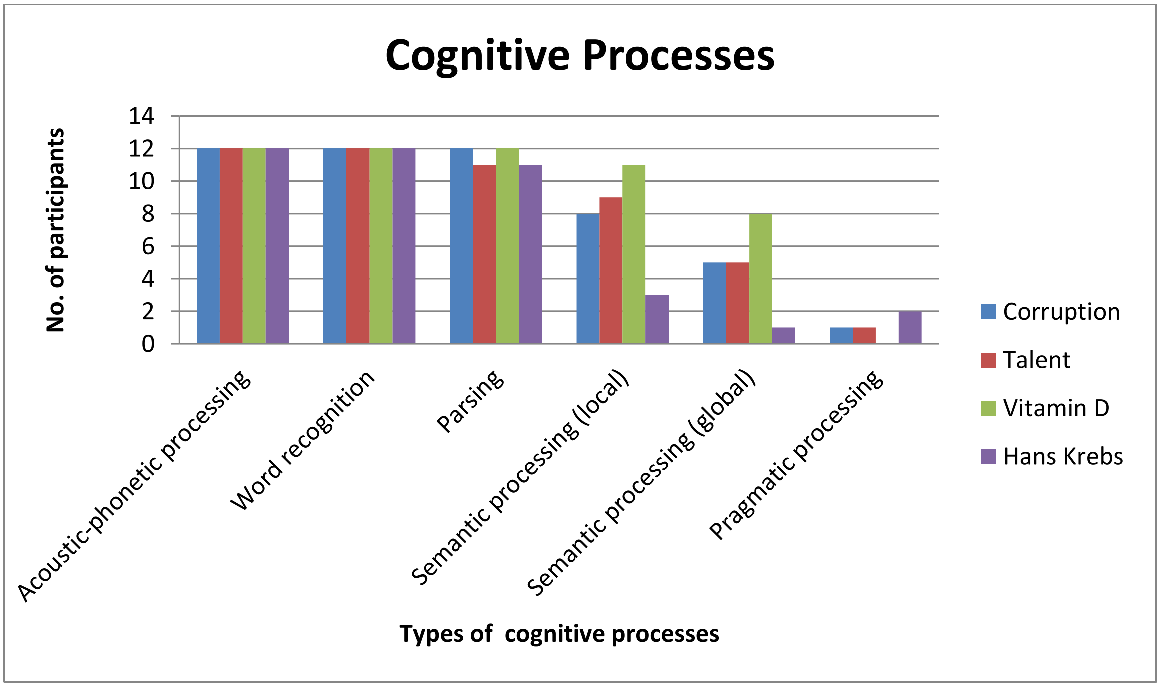


Figure 4.1: The participants' cognitive processes in understanding the listening input materials

As shown in Figure 4.1, a total of six cognitive processing types were identified in the data. The first three processes presented in the figure (from left to right) – acoustic-phonetic processing, word recognition, and parsing are categorized as linguistic processing and considered lower-level processes. The three processes to the right of the figure (semantic processing at the local level, semantic processing at the global level, and pragmatic processing) are comprehension processes, classed as higher-level processes.

As can be seen in the figure, all participants presented evidence that they engaged in the two lowest-level processes across the four listening topics, i.e., acoustic-phonetic processing and word recognition. Parsing was done by all participants while listening to *Corruption* and *Vitamin D* and only one participant did not provide observable evidence for parsing while listening to *Talent* and *Hans Krebs*. The number of participants appeared to engage in high-level processing, especially semantic processing at the discourse level and pragmatic processing, varied between the listening passages. Overall, high-level processes were relatively less used when compared to the low-level processes. Specifically, when listening to *Hans Krebs*, only a few participants (3 out of 12) presented evidence of activating semantic processing at the local level, only one participant engaged in semantic processing at the global level and two in pragmatic processing. In *Vitamin D*, although a large number of participants was found to engage in semantic processing (at a local level: 11 out of 12 participants, and at the global level: 8 out of 12), none of them engaged in pragmatic processing. About two-thirds of them activated semantic processing at the local level when listening to *Corruption* and *Talent*. Five of them engaged in semantic processing at the global level and only one person processed the text with pragmatic knowledge.

These findings suggest that fewer participants appeared to engage in higher-level processes such as semantic and pragmatic processing. In addition, they show that the cognitive processes activated may vary depending on the passages test-takers listen to. A more in-depth description of the participants' cognitive processing is provided in the following sections.

4.2.1 Acoustic-phonetic processing

Acoustic-phonetic processing occurs when a listener accesses acoustic sounds, registers the sounds, and converts the sounds into representations of the language's phonological system (see 2.4.7). At this stage of processing, phonemes or phonological forms which are the basic units of words are identified.

Acoustic-phonetic processing was one process automatically conducted by the participants in this study. None of the participants explicitly reported engaging in it. However, their activation of this process can be inferred from what they reported they heard or understood while listening. For instance, P1 explained that while listening to *Corruption* she heard 'public sector, estimated, and impact the poor '. From the words and phrases that the participants reported hearing and noted down while listening, it can be assumed that they all engaged in this process. In fact, acoustic-phonetic processing necessarily takes place prior to other higher processes, and only based on (part of) this process can other processes such as word recognition and semantic processing take place. Since the data obtained shows that the participants all processed the listening texts at levels higher than acoustic-phonetic processing, such as word recognition and parsing, it is possible to conclude that all participants engaged in acoustic-phonetic processing in order to understand the listening texts.

4.2.2 Word recognition

Word recognition is the process by which the listener segments continuous speech and identifies words or a series of words (phrases) in a speech stream (see 2.4.7). Word recognition is necessary in order to comprehend a listening text. This is because words/chunks of words are the basic units that convey or contribute to the meaning of utterances. In this study, it was found that all participants were able to recognize many words in the speech stream and wrote down the words they perceived as the key words in the passages. However, it is important to note that what constituted 'key words' varied between participants and that some of the words which the participants perceived as key could possibly not have been the key words in that particular text.

Word recognition is evidenced both in the stimulated recall data and participants' notes. For example, the following stimulated recall statement from P1 on *Corruption* shows that she recognised words/phrases:

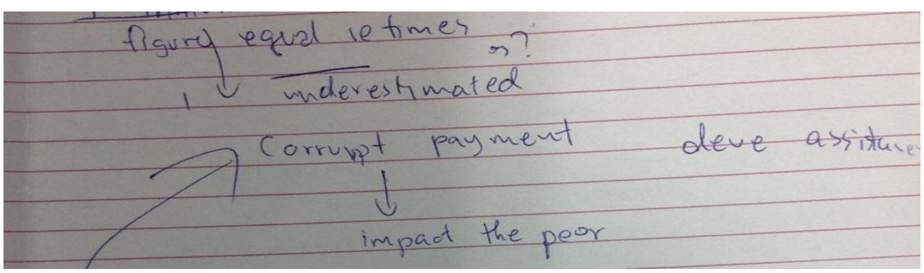
Quote 4.1

> I heard 'one trillion'…. I heard 'corruption', 'corrupt payment' and then I heard 'development assistant aids' (P1/Corruption).

In *Vitamin D*, P4 mentioned that he heard 'maintain blood calcium, bone, teeth, big trouble'. P8 reported that in his listening to *Hans Krebs*, he noted down some of the phrases he heard and thought were important, such as 'overcome all kinds of obstacles' and 'can't make silk purse'. In her listening to *Talent*, P6 indicated that she heard and wrote down key words such as 'analytical ability' and 'consultant management'.

The analysis of participants' notes confirms that they recognized words and phrases while listening. For example, Note 4.1 presents the words recognized and noted down by P1 when listening to *Corruption*.

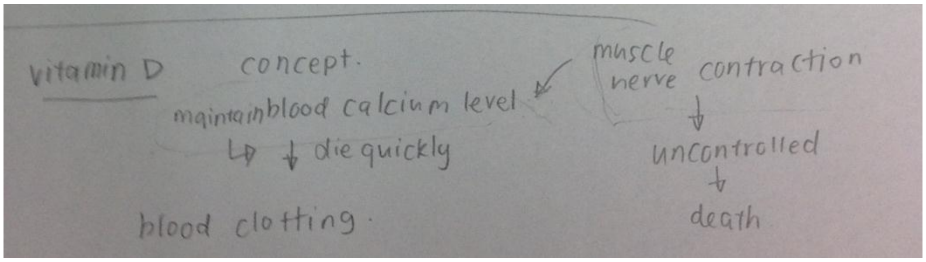
Note 4.1



(P1/*Corruption*)

P11 recorded the following in writing when listening to *Vitamin D*.

Note 4.2



(P11/*Vitamin D*)

4.2.3 Parsing

Parsing is the segmentation of a speech stream according to its syntactic or semantic cues to obtain a meaningful unit in a text. In the bottom-up processing view, parsing occurs when the listener combines words, and maps them onto the syntactic and/or semantic structures of the language. The top-down processing view maintains that parsing occurs when the listener relies on context to identify meaningful text units. The result of the parsing process is propositions which generally consist of one predicate and one argument (an agent and a predicate which includes an object or a verb modifier) (Anderson, 1985; Kintsch & van Dijk, 1978).

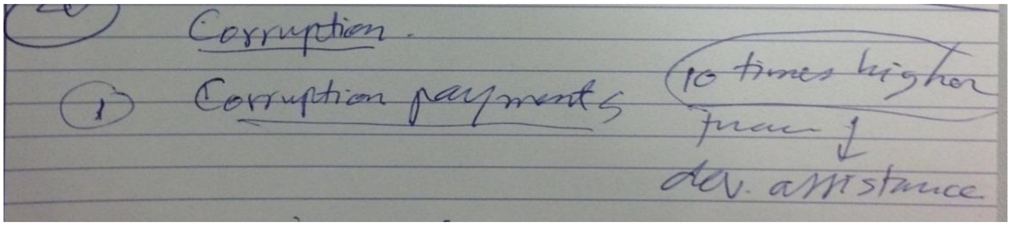
The analysis of the stimulated recall data, the participants' notes, and the summary content shows that parsing is one of the processes activated by every participant, but with differences in numbers when compared across listening passages. All participants appeared to have engaged in parsing when listening to *Corruption* and *Vitamin D*. In their listening to *Talent* and *Hans Krebs*, 11 participants were shown to have engaged in this process. The participant who did not show evidence of parsing these texts is P12 in his listening to *Talent* and *Hans Krebs*.

In the stimulated recalls, when asked what they were listening or paying attention to, the participants said out loud the propositions that they could parse from the listening passages, some of which were not grammatically correct but meaningful to the participants. Examples of the meaningful units segmented and reported by the participants are as follows.

> ‘people need assistance’ (P1/*Corruption*)
>
> ‘it’s very difficult to capture the number’ (P6/*Corruption*)
>
> ‘he is the wonderful example’ and ‘he can overcome all kinds of obstacles’ (P11/*Hans Krebs*)
>
> ‘he published research paper’ and ‘he is one of the great scientist’. (P4/*Hans Krebs*)
>
> ‘talent is defined in different ways’ (P1/*Talent*)
>
> ‘some define talent in a restrictive way’ (P12/*Talent*)
>
> ‘Vitamin D maintains blood calcium’ (P9/*Vitamin D*)

In addition to the verbal report data, the notes taken by the participants during the listening tasks confirmed, to some extent, that the participants engaged in parsing. The following are examples of notes which show the results of parsing processes. P4 explained that he heard, ‘corrupt payments is ten time higher than development assistant aid’ (Note 4.3). In his notes, although he did not put the words together as a complete sentence, he conveyed the meaning through an arrow and the specific words he wrote down.

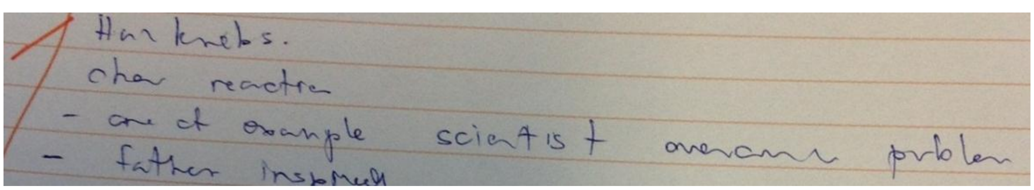
Note 4.3



(P4/*Corruption*)

Note 4.4 is another example of a participant engaging in parsing. In this note, the participant (P10) explained that she wrote down what she heard and considered to be the main points of the passage when listening to *Hans Krebs*. As explained by the participant, she understood that the listening was about Hans Krebs, so she wrote Hans Krebs on the first line. Then, on the second line, she wrote ‘chem reaction’ because she understood that Hans Krebs studied chemical reactions. Below this, she wrote ‘one of example/ scientist/ overcome problem’ because of her understanding that Hans Krebs is a scientist who has overcome problems. From the participant’s explanation of the notes she took and her actual notes, it is, thus, possible to conclude that this participant engaged in parsing.

Note 4.4



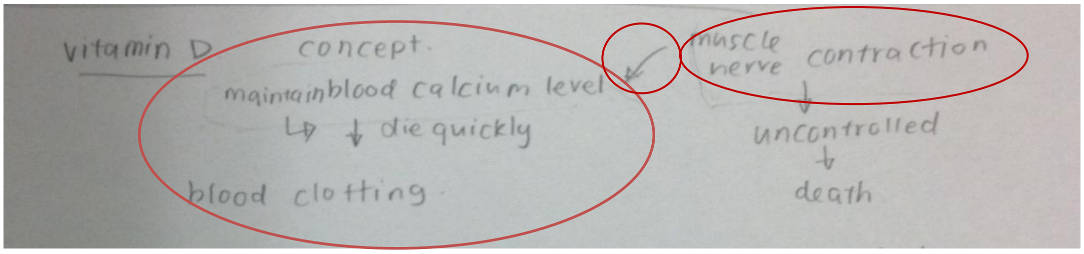
(P10/*Hans Krebs*)

4.2.4 Semantic processing at the local level

For listeners to understand the meaning of texts beyond their literal meaning, they have to engage in semantic processing which takes place when listeners relate chunks of information together and link these to their world knowledge. Semantic processing principally operates at two different levels, i.e. local and global. At the local level, semantic processing enables the listener to make connections between individual propositions to establish propositional meaning. In this study, as shown in Figure 4.1, fewer participants appeared to engage in semantic processing at the local level compared to those engaging in low-level processes (acoustic-phonetic processing, word recognition, and parsing).

Compared across the four listening passages, a large number of the participants employed semantic processing at the local level (11 out of 12) when listening to *Vitamin D*. 9 of them activated it when listening to *Talent*, 8 when listening to *Corruption*, and only 3 when listening to *Hans Krebs*. Semantic processing at this level was evident in the participants' stimulated recall and also in the notes and content summaries. For example, when listening to *Vitamin D*, P11 described that she knew Vitamin D maintains calcium level and it relates to nerve contraction. In her note, presented below, she used an arrow to link these two pieces of information together.

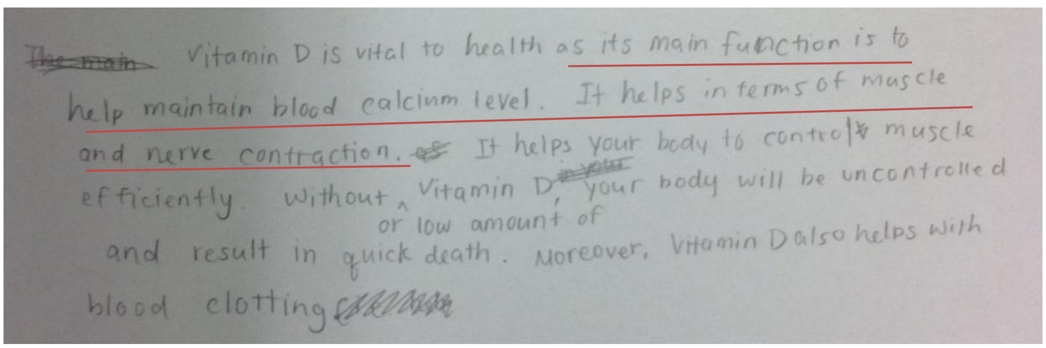
Note 4.5



(P11/*Vitamin D*)

In addition, in her written summary, she stated:

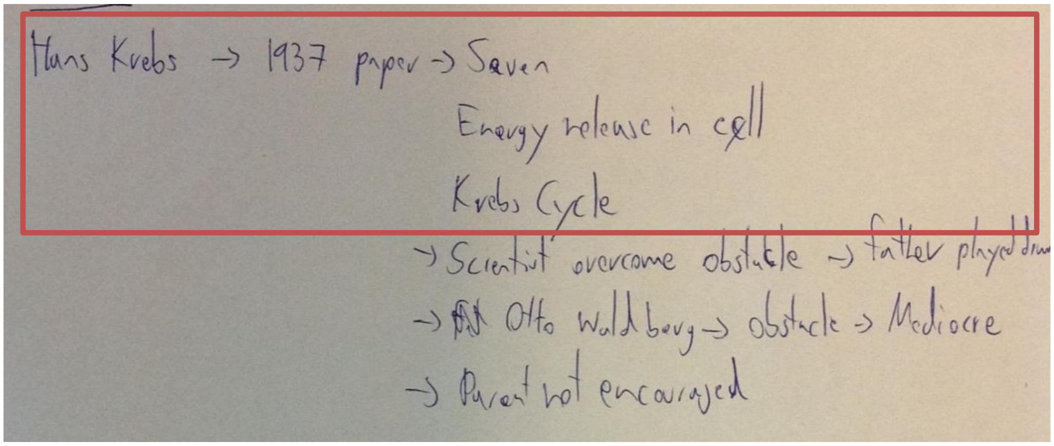
Written Summary 4.1


~~The main~~ Vitamin D is vital to health as its main function is to help maintain blood calcium level. It helps in terms of muscle and nerve contraction. It helps your body to control muscle efficiently. without or low amount of Vitamin D, ~~your~~ your body will be uncontrolled and result in quick death. Moreover, Vitamin D also helps with blood clotting

(P11/*Vitamin D*)

Another participant who appeared to have engaged in semantic processing at the local level is P3. Although in his notes (Note 4.6), the participant did not include complete sentences, his stimulated recall data showed that he understood several single ideas in the text, including Hans Krebs was famous, and Hans Krebs published a paper about energy release in cell.

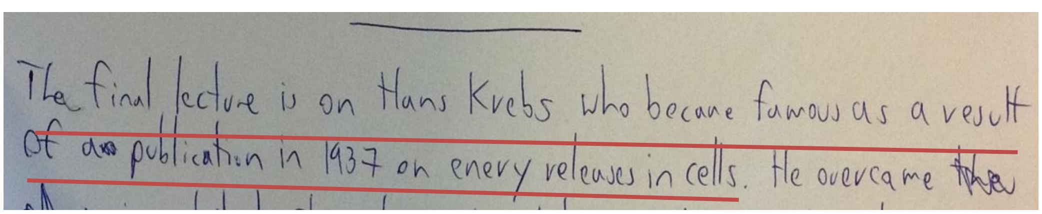
Note 4.6


Hans Krebs → 1937 paper → Saven
Energy release in cell
Krebs Cycle
→ Scientist overcome obstacle → father played dim
→ ~~At~~ Otto Waldberg → obstacle → Mediocre
→ Parent not encouraged

(P3/*Hans Krebs*)

From the note, the participant then wrote the content summary 4.1. The beginning of the summary content shows that the participant linked ‘1937 paper’ and ‘energy release in cell’ together as one set of information or one single idea of the listening. It can thus be inferred that this participant engaged in semantic processing at the local level.

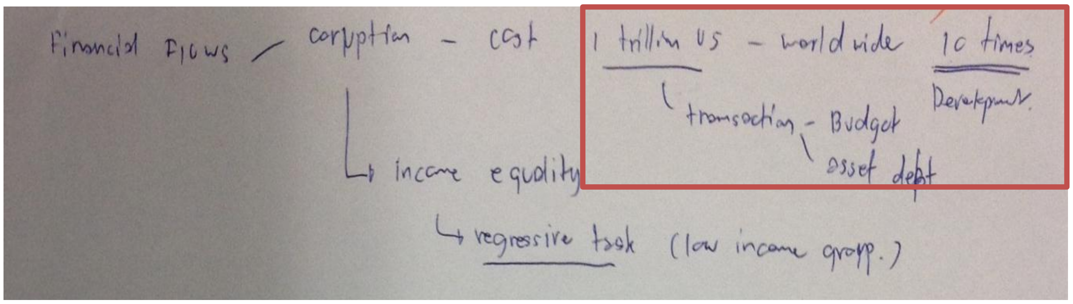
Written Summary 4.2


The final lecture is on Hans Krebs who became famous as a result of ~~a~~ publication in 1937 on enery releases in cells. He overcame ~~the~~

(P3/*Hans Krebs*)

When listening to *Corruption*, a large number of participants (8 out of 12) appeared to have engaged in semantic processing at the local level. P12, for example, reported in his stimulated recall that he understood ‘corruption is the problem, corruption is about one trillion US dollars, and corruption relates to transaction and budget asset debt'. In his notes, as can been seen in the rectangle in Note 4.7, this participant visually linked these pieces of information together.

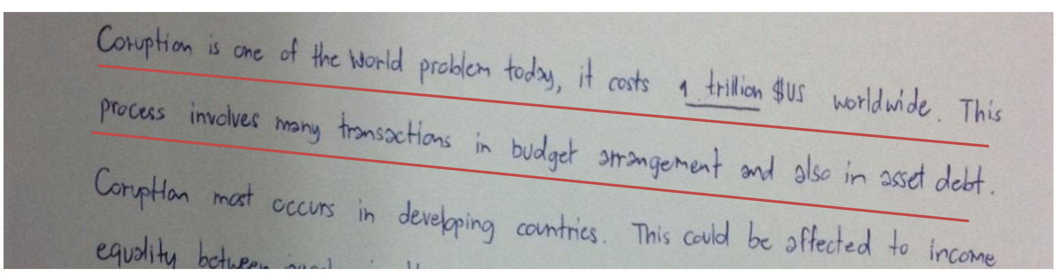
Note 4.7



(P12/*Corruption*)

This participant's written summary (4.2) also confirms that he engaged in semantic processing at the local level as he appeared to link these pieces of information together and included it in his summary:

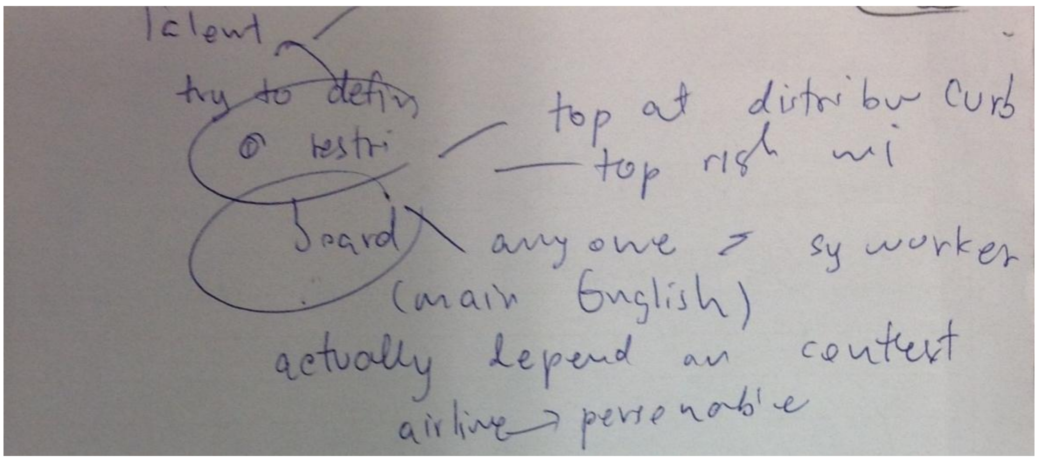
Written Summary 4.3



(P12/*Corruption*)

When listening to *Talent*, nine participants appeared to have engaged in semantic processing at the local level. P1, for example, described in her stimulated recall that she knew that there are two different ways to define talent: a restrictive way and a broad way. She recorded these words in her notes (4.9) and circled the two words because she considered them to be key points in the listening and each of the words has their own description. From the two key words, she then drew short lines to link them to their description.

Note 4.8



(P1/*Talent*)

The transcript of her oral summary (4.1) also shows that this participant engaged in semantic processing at the local level. That is, she connected the individual pieces of information as recorded in her notes and put them together in complete sentences.

Oral Summary 4.1

> 'This is about the word talent and the word talent is defined in different ways. There are actually two ways that people define talent. The first way is quite restricted and for this it means like people at the top of the distribution curve ……' (P1/*Talent*)

4.2.5 Semantic processing at the global level

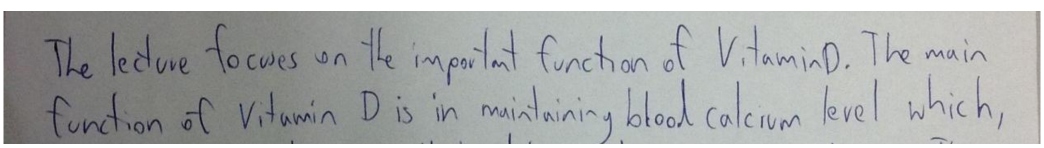
At the global level of semantic processing or at the discourse level, listeners associate concepts with linguistic segments to determine the linguistic element or concept that remains as the focus of the passage they have listened to. The listeners have to make use of the linguistic segments obtained from the previous processing stages and their discourse knowledge to determine the main point of the listening. Some linguistic cues that are helpful include connectives such as *because, and, but, so, for* that help provide relational information between individual concepts, syntactic cues that indicate the relations between subjects and predicates, and discourse-level (genre) cues that help to outline the organization of the information and the way the information is presented.

The analysis of the stimulated recall transcripts, the notes, and the summary content shows that the number of participants who engaged in semantic processing at the global level was lower than those engaging in the local level of semantic processing. Comparing across the four listening topics, the highest number of participants (8 out of 12) performing this process was found for *Vitamin D*. 5 out of 12 participants activated it when listening to *Corruption* and *Talent*, and only one participant activated this when listening to *Hans Krebs*. The main indication of participants' successful performance of semantic processing at the global or discourse level was an accurate main point provided in their summary content. Below, examples of the activation of this process are presented for each listening topic.

The highest number of participants activated semantic processing at the global level while listening to *Vitamin D*. Several points are mentioned in this passage, including the central concept of Vitamin D, the function that Vitamin D has by accident, the real function of Vitamin D, and the effect of a lack of Vitamin D. The participants who appeared to have successfully engaged in semantic processing at the discourse level are those who interconnected individual concepts and linked them to the theme of the topic, whilst at the same time maintaining their focus on the main point (the real function or advantage of Vitamin D is maintaining blood calcium within a narrow range). Those participants who kept their focus on this theme and included it in their summary are considered to have performed this process successfully. Below are examples of the summary content from these participants.

> *Oral Summary 4.2*
>
> Ok, umm the recording talks about the Vitamin D which is about the central concept of Vitamin D. It said that umm it controls blood calcium. The benefit of Vitamin D is to control… (P8/*Vitamin D*)

> *Written Summary 4.4*


The lecture focuses on the important function of VitaminD. The main function of Vitamin D is in maintaining blood calcium level which,

(P3/*Vitamin D*)

The stimulated recall data also revealed that some participants engaged in this process. This can be inferred from the participants' verbalization of their understanding of the listening passage, which was relevant to the main point of the passage. P6, for instance, mentioned:

Quote 4.2

> I know that the main focus was about the main function of Vitamin D that helps maintain blood calcium level. It's not about maintaining bones and teeth as generally understood. (P6/*Vitamin D*)

This reflection then is supported by the participant's oral task summary:

Oral Summary 4.3

> This lecture talk about function of Vitamin D and its central concept which is very important for blood calcium… (P6/*Vitamin D*)

When listening to *Talent*, fewer participants (5 out of 12) appear to have engaged in semantic processing at the global or discourse level. Different concepts are provided in this passage, including broad and narrow definitions of talent, the meaning of talent in different businesses (e.g., in management consultancy firms and in airline businesses), and the meaning of talent according to the speaker. However, the gist of this listening passage is that talent can be differently defined depending on the context. The participants who included this point in their summary were those considered successfully engaging in this process. Below are examples of the summary content provided by successful participants.

Oral Summary 4.4

> Ok, the second lecture discuss the idea of talent in that it has been defined in many different ways. If whether you know from the alpha worker or to anybody in the workforce, but what the…. (P3/*Talent*)

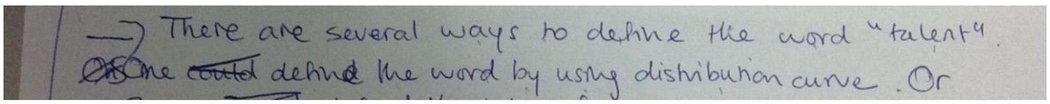
Oral Summary 4.5

> This is about the word talent and the word talent is defined in different ways. There are actually two ways that people define talent. The first way is quite restricted and for this it means like… (P1/*Talent*)

Oral Summary 4.6

Umm what is talent? Well, it is a very difficult to answer the question because err anyone has a different ways to think about the talent. Err in some company, talent might mean err top… (P10/*Talent*)

Written Summary 4.5


→ There are several ways to define the word "talent".
One ~~could~~ define the word by using distribution curve. Or

(P8/*Talent*)

With reference to *Corruption*, five participants engaged successfully in semantic processing at the global or discourse level. Points which are provided in this listening passage include the economic and social cost created by corruption, a large figure of one trillion U.S. dollars paid in bribes, a comparison of the money paid for corruption versus international development assistance, the impact of corruption on the poor, and a regressive tax on household incomes. However, the main message of the passage is that corruption is a social problem and it affects the poor more than the rich. One of the participants who successfully performed this process is P3. In the stimulated recall, he stated:

Quote 4.3

I knew that the passage was about corruption. There were several details here. I was thinking what actually the main point was… I was thinking that the speaker wanted to point out the impact of corruption on social service…corruption affected the poor more because they had to pay bribes.. (P3/*Corruption*)

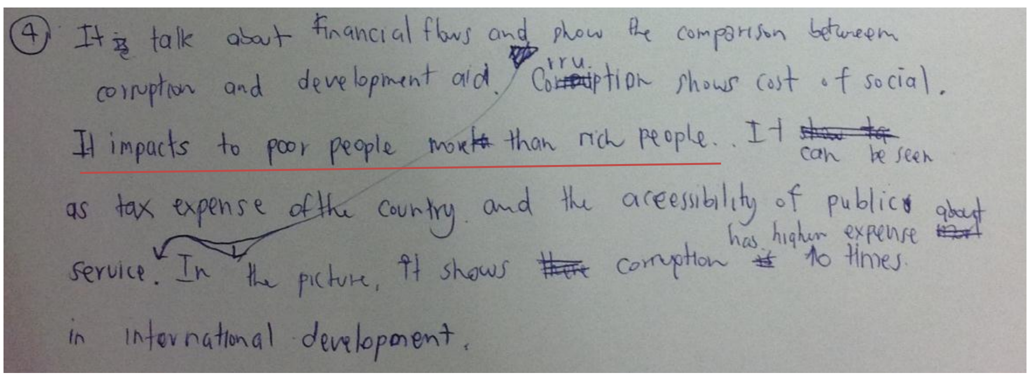
In his oral summary, this participant presented:

Oral Summary 4.7

> The …ah lecture describe the bad effect of bribery in that in poor country where bribery takes place into the amount of one trillion dollar. This corruption has ...err.. very bad effect on the poor people in that they need to pay bribery to access to the amenity that they require but because they don't have such money to pay for the bribery they become worse out of it and instead of giving assistance they had to err.., you know, to pay up for the bribe. (P3/*Corruption*)

Written Summary 4.6 is an example from another participant (P5) who was able to correctly identify the main point of the text. She included in her summary that corruption impacts the poor more than the rich.

Written Summary 4.6

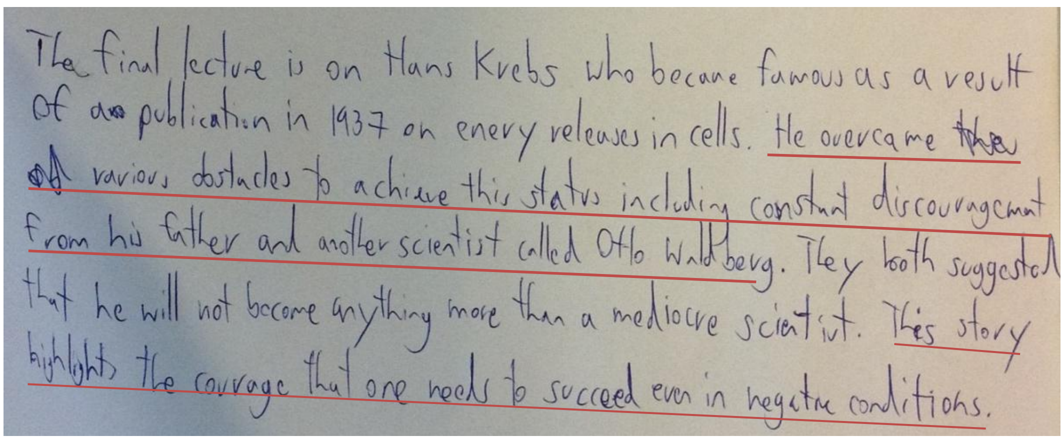

(4) It ~~is~~ talk about financial flows and show the comparison between corruption and development aid. Corruption shows cost of social. It impacts to poor people more than rich people. It ~~show to~~ can be seen as tax expense of the country and the accessibility of public service. In the picture, it shows corruption has higher expense 10 times about in international development.

(P5/*Corruption*)

In *Hans Krebs*, where the main point is less explicit compared to *Vitamin D* and *Talent*, only one participant appeared to have successfully engaged in semantic processing at the discourse level. The listening began by describing a person whose name is Hans Krebs, followed by his publications, and his great success as a chemist. Then the listening shifts to describing how people – in this context referring to Hans Krebs – can overcome obstacles despite a lack of parental support. The data analysis shows that most of the participants in this study mistakenly understood that the listening was mainly about Hans

Krebs and how he became a famous chemist. Although the participants appeared to understand single concepts mentioned in the listening, they did not successfully identify the main point of the listening text, namely that Hans Krebs is a wonderful example of a human being who could overcome obstacles and become successful in life. Below is the content summary of the only participant (P3) who successfully processed the text at a discourse level and accurately provided the main point.

Written Summary 4.7


The final lecture is on Hans Krebs who became famous as a result of ~~a~~ publication in 1937 on enery releases in cells. He overcame ~~the~~ ~~of~~ various obstacles to achieve this status including constant discouragement from his father and another scientist called Otto Waldberg. They both suggested that he will not become anything more than a mediocre scientist. This story highlights the courage that one needs to succeed even in negative conditions.

(P3/*Hans Krebs*)

This participant's stimulated recall data also show that he engaged in processing at the discourse level:

Quote 4.4

I understood that the main point of the listening was on the second half of the listening that explained how Hans Krebs became successful though this father and his teacher said he would never be a great scientist, not the first part that focuses mainly on his work. (P3/*Hans Krebs*)

4.2.6 Pragmatic processing

Often, the true meaning of a text is not explicitly stated. Thus, to understand its intended meaning, listeners have to use their pragmatic knowledge to determine the speaker's intentions. That is, they need to elaborate on the linguistic information and bring in their social and cultural knowledge about the context of communication. This elaboration is referred to as pragmatic processing. Pragmatic processing is thus a type of processing which assists the listener in understanding the meaning that is left unsaid or implied by the speaker.

In the present study, pragmatic processing, compared to the other cognitive processes, was performed by the smallest number of the participants (only two). This could be due to two reasons. First, most of the ideas in the texts were communicated explicitly, and these ideas, to some participants, might be enough to produce a summary of the listening. It was thus unlikely to be necessary for them to rely on contextual and/or extra-textual information to understand the text. A second reason may be related to a participant's ability to process the text. As revealed in the stimulated data, some participants struggled with processing the text at lower-level processes such as parsing and did not manage to go beyond these processes to understand the actual meaning of the text.

With reference to *Corruption*, the analysis of the stimulated data shows that one participant engaged in pragmatic processing. This participant (P3) relied on pragmatic processing to identify the context of the passage, which he thought was unclear. He explained in his stimulated recall:

Quote 4.5

> …here (*Corruption*), the speaker didn't make it clear what the context was when he talked about social issue and what he wanted to achieve by saying that. But, the tone of the lecture was serious. I guess he was talking about corruption problems or an impact of corruption on the poor. (P3/*Corruption*)

The same participant used the tone of the lecture and the visual information provided to understand that the speaker wanted to point out corruption problems:

Quote 4.6

> I knew that the speaker wanted to point out to problems related to corruption. I saw the numbers given in the graph….The tone of the lecture which was quite serious, more serious than the other lectures (*Talent* and *Vitamin D*). (P3/*Corruption*)

The same participant (P3) also appeared to activate pragmatic processing when listening to *Talent* and *Hans Krebs*. While listening to *Talent* in particular, he analysed the language used in the lecture together with the image provided to determine the key message the speaker wanted to deliver to the listeners. He explained:

Quote 4.7

> This lecture (*Talent*) was easier than the first lecture (*Corruption*). The language was simpler and more straightforward and picture was very much related to the listening passage….The speaker related the concepts of talent to human pyramid from the picture given. ….I understand that the person who is on the top of everything is a talented person. (P3/*Talent*)

While listening to *Hans Krebs*, this participant tried to identify the context of the lecture or which study area this lecture would relate to, i.e. science or psychology. He analysed the visual information and compared it against the linguistic information he gained from the listening. Then he realized that the picture did not well describe the main point the speaker wanted to deliver, which was ‘the person can overcome their obstacles no matter what we say to them’. The participant described:

Quote 4.8

> This sounded like pure science at the beginning but actually it was not. It was more towards a psychology lecture. It was a bit confusing at the beginning to know what the main point was. The photo given did not really match the listening passage, really not. I was wondering why the lecturer started with his great success because the main point was not that. (P3/*Hans Krebs*)

While listening to the same text (*Hans Krebs*), another participant (P7) also appeared to engage in pragmatic processing. This participant was, however, not as successful as P3. The participant relied on the tone of the listening passage to determine the speaker’s purpose. However, to him the listening text did not sound academic and he did not consider it to be a proper lecture. He thus misinterpreted the speaker’s intention in delivery of the message and misunderstood the text’s main point.

Quote 4.9

> I thought a lecture had to be given in a more serious tone. This listening (*Hans Krebs*), to me sounded more like storytelling, so I just followed the story. I thought the speaker of this listening passage wanted to describe the life of a famous scientist. (P7/*Hans Krebs*)

No evidence was found of participants engaging in pragmatic processing when listening to *Vitamin D*. This could be because the passage was straightforward and explicit. This seems to be plausible, as some participants obtained a full mark for the content of their summary without having performed (or showing any evidence of) pragmatic processing.

4.3 Summary

A total of six cognitive processes (acoustic-phonetic processing, word recognition, parsing, semantic processing at the local level and the global or discourse level, and pragmatic processing) have been shown to have been activated by the participants when trying to comprehend the listening inputs of the listening-to-summarize tasks investigated in this study. Although every participant showed evidence of the lower-level processes of acoustic-phonetic processing, word recognition, and parsing, a small number appeared to have engaged successfully in high-level processes (semantic processes at the local and the discourse levels, and pragmatic processes). Many participants were found to have difficulties with comprehension processing. One listening difficulty that the participants encountered, especially when listening to *Corruption, Talent,* and *Hans Krebs,* was semantic processing at the global level or associating individual concepts with the theme of the passages in order to understand the texts' discourse meaning. About half of the participants appeared to have attempted to understand the main point of the texts. However, their attempt did not appear to be successful. This was partly because their text processing at the lower-level (linguistic processing) was not as successful as it should be. For example, they were not able to recognize some words which conveyed the key meaning of the texts or they thought that the words they obtained were key words of the text, which in fact they were not.

Chapter 5 Cognitive strategies

5.1 Introduction

This chapter and the next (Chapter 6) present the findings of the participants' use of strategies to enhance their comprehension of the listening-to-summarize input texts. It has been suggested in the literature that two categories of strategies are used for this purpose, i.e. cognitive and metacognitive strategies. Cognitive strategies are those strategies language users employ to solve problems occurring during the comprehension process (Bachman & Palmer, 2010). Metacognitive strategies, on the other hand, are used to oversee and manage language use including the use of cognitive strategies (Bachman & Palmer, 2010). This chapter provides the findings on the cognitive strategies the participants used to facilitate their cognitive processing (presented in Chapter 4) and to solve any comprehension problems they had whilst completing the listening-to-summarize tasks. The chapter begins by presenting an overall picture of the cognitive strategies used and the number of participants who used each of them, followed by a detailed coverage of each strategy used in separate sections. The next chapter (6) will describe the participants' use of metacognitive strategies.

5.2 Cognitive Strategies

Gaps in listening comprehension occasionally take place, and they are even more likely in the case of L2 listening, as compared to L1, due to L2 listeners' more limited linguistic knowledge and less automatized language processing (Buck, 2001; Goh, 2002; Rost, 2011). Taking Anderson's (1985) three stages of comprehension processing, i.e. perceptual processing, parsing, and utilization, as a framework, Goh (2002) explored and described listening difficulties/gaps at different stages of processing. Common problems at the perceptual processing stage, where the listeners decode sounds they hear and group them into words or series of words, are: 1) being able to recognize words they know but not their meaning, 2) neglecting the next part whilst thinking about meaning, 3) not being able to chunk streams of speech, 4) missing the beginning of sentences, and 5) either concentrating too hard or being unable to concentrate (Goh, 2002). Problems revealed at the parsing stage, where the words are transformed into a mental representation of the combined meaning of the words, include: 1) quickly forgetting what is heard, 2) not being able to reconstruct the gist from the words the listener manages to hear, and 3) not being able to understand subsequent parts of the input because of earlier problems. At the utilization stage, where the discourse meaning of the passage is obtained, the following problems have been identified: 1) not being able to understand the message despite understanding every word, and 2) being confused about the key ideas in the message (Goh, 2002).

In the present study, a number of listening problems and difficulties were also encountered by the participants. To diminish these problems and enhance listening comprehension, the participants activated a number of strategies, including the cognitive strategies which are the focus of this chapter. Following Vandergrift and Goh (2012), the use of strategies is considered available to the user's consciousness. Therefore, the participants' verbal reports of what they were doing while listening were used as the main data source to investigate the strategies used. The participants' notes and their summary content were also analyzed to supplement the stimulated recall data.

The analyses of the recalls, notes, and summaries showed that the participants applied a number of strategies in order to understand the four listening input materials (*Corruption*, *Talent*, *Vitamin D*, and *Hans Krebs*), with the aim of summarizing these either in oral or written form (see Figure 5.1).

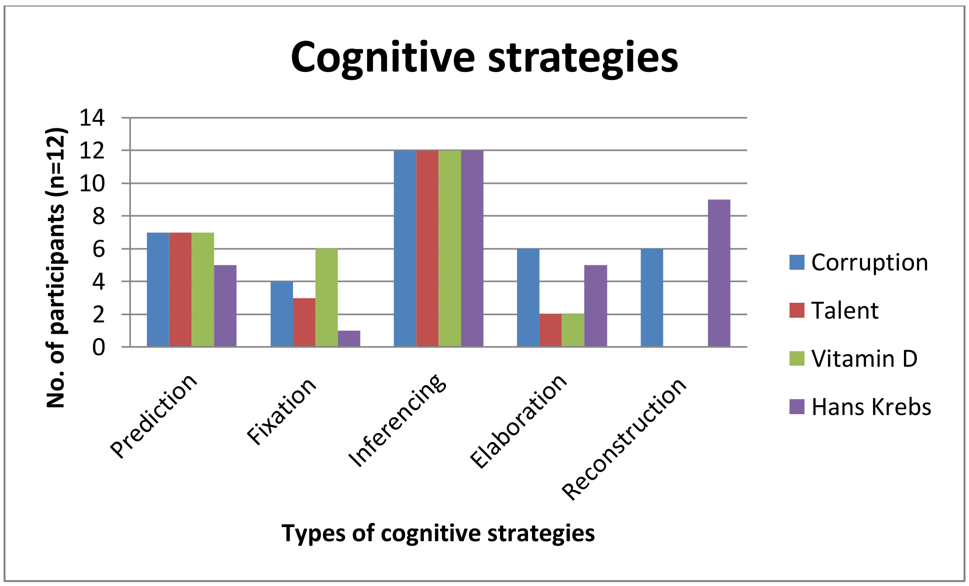


Figure 5.1: Cognitive strategies used by the participants to solve listening problems

Figure 5.1 presents the cognitive strategies and the number of participants employing them to understand each listening passage. The strategies are ordered according to their occurrence during the listening tasks. Starting from the left of the X-axis is the prediction strategy, occurring at the beginning of the listening tasks. Three cognitive strategies activated after that were fixation, inferencing, and elaboration. After these three strategies and towards the end of the listening some used reconstruction strategies. Among these five strategies, inferencing and prediction appear to be ‘more popular’ than the others. All participants reported making inferences when listening to all four listening passages. Regarding the use of prediction, a similar number of the participants (7 out of 12), but different individuals, used it when listening to *Corruption*, *Talent* and *Vitamin D*. Only five participants reported predicting the listening content when listening to *Hans Krebs*. The elaboration and reconstruction strategies were used by different numbers of the participants, varying between the listening topics. Higher numbers of participants used these two strategies when listening to *Corruption* and *Hans Krebs* than when listening to *Talent* and *Vitamin D*. In fact, no participants appeared to use reconstruction when listening to *Talent* and *Vitamin D* at all. Fixation was a strategy used by a small number of the participants, especially in *Hans Krebs*, where only one participant reported using it. A more detailed report of the use of each strategy is provided below, supported by quotes from the stimulated recalls, the participants’ notes, and their summary content where relevant.

5.2.1 Prediction

Prediction or forward inferencing, as referred to by Anderson (1985), takes place when the listeners activate their schemata related to the listening topic and combine it with the information they have gained from the listening to anticipate what is coming next in the listening passage. Prediction occurs at two levels, i.e. the global and local levels. Predicting globally enables the listener to predict the general content of the text, whereas at the local level the listener anticipates details of the text or the immediately upcoming part of the text such as words/phrases or an idea.

Compared to the other strategies used, the analysis showed that prediction is the second most prevalent strategy, following after inferencing. Seven participants (out of 12), but a different combination of individuals for each input text, used this strategy when listening to *Corruption*, *Talent* and *Vitamin D*. Five participants applied this strategy when listening to *Hans Krebs*. Predictions were mainly made at the global level; that is, to anticipate the main point the speaker wanted to deliver in each of the listening passages. Prediction at the global level took place predominantly in the beginning of the listening task after the participants had seen the image and listened to the first few sentences. Interestingly, similar predictions for each listening topic were made across the participants. In *Hans Krebs*, the participants predicted that the listening passage would be about the biography of the man pictured on the screen. They, for example, reported:

Quote 5.1

I saw the picture of a man. I guessed it was going to be about a biography of this man so I wrote down 'biography'. (P2/*Hans Krebs*)

Quote 5.2

I looked at the picture and the words under it. They were 'career, in career'. I thought the story was about this man and what he had done. I guessed he was either a historian or a scientist. (P5/*Hans Krebs*)

Quote 5.3

I saw the picture with the name under it. I thought the passage was about this man, Hans Krebs, and he must be an expert in science. I guessed the story would be about his professional career. The photo showed the peak of his career. I thought he was famous. (P8/*Hans Krebs*)

Quote 5.4

I thought the speaker was going to talk about a biography of this man (Hans Krebs). I saw the words 'peak of his career' under his photo. (P11/*Hans Krebs*)

In *Corruption,* where a bar chart was given, the participants made the global prediction that the listening passage was going to make a comparison between the two bars, i.e. between corrupt payment and development assistance. They, for example, said:

> *Quote 5.5*
>
> I saw the graph and heard that it was ten times. I was thinking that it would be about the corruption that is 10 times higher than assistant development. (P2/*Corruption*)
>
> *Quote 5.6*
>
> I saw two bars in the graph and the words 'corruption' and 'development assistant' under each of them. I anticipated that the listening text was going to compare between corruption and development assistance. (P9/*Corruption*)

In *Talent*, after hearing that talent can be defined in different ways, the participants globally predicted that the listening passage was going to provide different definitions of talent. Some of them indicated:

> *Quote 5.7*
>
> I heard 'there are different ways to define talent'. Then I predicted that I was going to hear the first, the second, and the third... ways of defining talent. (P1/*Talent*)
>
> *Quote 5.8*
>
> I looked at the screen and I heard it was not easy to define talent. I was thinking that later in this passage, different people would give their own definitions of talent. (P2/*Talent*)
>
> *Quote 5.9*
>
> I heard "What is talent?". I thought it told what the story was going to be about. I predicted it was about the meaning of talent. (P11/*Talent*)

In a similar manner, in *Vitamin D*, the participants made the global prediction, after listening to the first few sentences, that the story was about the benefits of Vitamin D. As they described:

> *Quote 5.10*
>
> I saw Vitamin D on the screen. Then I heard 'Vitamin D and central function', I knew the story was going to be about the function of Vitamin D in our body system. (P6/*Vitamin D*)
>
> *Quote 5.11*
>
> After listening to the first sentence, I predicted that it was either about the advantages or disadvantages of Vitamin D, or both. (P8/*Vitamin D*)
>
> *Quote 5.12*
>
> The first thing I was thinking of when hearing 'the function of Vitamin D' is the benefits of it to bones and teeth. (P3/*Vitamin D*)

Predictions at the local level, where details of the texts are predicted to be heard, were also made by the participants. Two participants anticipated a more specific detail of the listening. When listening to *Corruption*, one participant described:

> *Quote 5.13*
>
> From the graph I saw and the beginning of text I listened to, I knew that corruption has an impact on a social process and it affects the poor. I predicted that some numbers would be given by the speaker to explain the graphic information. (P3/*Corruption*)

Another participant anticipated that the upcoming text would be about muscles and bones after hearing 'blood calcium' in *Vitamin D*.

Quote 5.14

> When I heard 'blood calcium', I thought how calcium works in our body system. Then I predicted that it helps to maintain strong bones. (P4/*Vitamin D*)

Based on the beginning of the text and their prior knowledge, the participants predicted what they were going to hear next. However, these predictions were not always accurate, as shown in Quotes 5.1-5.4 on *Hans Krebs*, where the predictions were not in line with the text. The text's main point was that people can become successful even when they lack parental support. The participants, however, anticipated hearing the story of a famous person who in fact was simply mentioned as an example. What they predicted as the main point was only a piece of the details.

Since prediction occurred mainly in the beginning of the listening tasks and was based only on partial information, prediction could thus go wrong. Other metacognitive strategies (see Chapter 6) such as comprehension monitoring (verifying prediction) need to be activated to make prediction more effective.

5.2.2 Fixation

Goh (2002) described that fixation is when listeners focus their attention in order to understand a small part of a spoken text. They, for instance, pause to 1) think about the spelling of unfamiliar words, 2) think about the meaning of words or parts of the text, 3) memorize or repeat the sounds of unfamiliar words, and 4) memorize words or phrases for later processing. The analysis of the stimulated recalls showed that fixation was used mainly at the perceptual stage or linguistic processing level, especially when the listeners tried to recognize words in a speech stream. The highest number of participants (6 out of

12) using fixation was found for *Vitamin D*. Four participants reported using it when listening to *Corruption,* three for *Talent*, and only for *Hans Krebs*.

The data revealed that fixation was used at different levels of processing and for different purposes. At the linguistic processing stage, fixation was used to deal with lexical difficulties. One main goal was to identify unfamiliar or unknown words in the speech stream. Two participants, for instance, indicated that they stopped to think about the word 'brain' while listening to *Talent*. However, neither of them was successful in this manner because they did not manage to decode the right word, 'brain', as revealed in their stimulated recall data:

Quote 5.15

> I stopped. I didn't know the word I had just heard. It's something like brib...power. I didn't know what that was. (P4/*Talent*)

Quote 5.16

> I was thinking that there were several key points in this listening passage, too many ideas to memorize, too. I couldn't get them all. Here it sounded like 'brie..' but I didn't know what it was. Was it 'bright'? (P6/*Talent*)

For the same passage, another participant (P5) stopped to think about the word 'different'. However, she was not successful as she indicated in her verbal report that she was not sure what the word was. To her, it could be 'difficulty', which is wrong because the original word was 'different'.

Quote 5.17

> I didn't know the word. I knew before that it was 'definition', but the later word was not clear to me. I stopped and wondered whether it was 'difficulty'. I didn't know. (P5/*Talent*)

Some participants, although they were able to identify the words they heard in the continuous speech, did not remember their spelling. Therefore, they used fixation or stopped to figure out the spelling of the words they had just heard. Participant 1, for example, reported:

> *Quote 5.18*
>
> I have to admit that my spelling skill is so bad. Here, I stopped to think about the spelling of the word 'death'. I was trying to use the sounds to find it out. (P1/*Vitamin D*)

One of the words that received a lot of attention from the participants, and for which some paused to figure out its spelling, is ‘t-e-t-a-n-y’, a term used in *Vitamin D* to refer to the disease caused by a lack of Vitamin D in blood calcium.

> *Quote 5.19*
>
> I heard 'tetany'. I knew it was about a disease, but I didn't know how to spell it out. I stopped and tried to find it out from the sound I heard. Was it ‘t-e-s-t-i-n-y’? (P5/*Vitamin D*)
>
> *Quote 5.20*
>
> I knew that it was the name of a disease, but I'm not familiar with it. I stopped to think about the spelling of the word. I guessed it is 't-e-t-i-n-y'. I’m not sure whether I got it right. (P3/*Vitamin D*)

It is not entirely clear why the participants wanted to figure out the spelling of the words while they were listening for comprehension. Some participants (e.g., P3 and P12) were trying to obtain the word spelling in order to use it correctly in their written summary of *Vitamin D*. However, others (e.g., P5) stopped to think about word meaning in the task, where they were required to orally summarize the listening. In this latter case, it could be because the participant was able to understand the likely meaning of the word (tetany) and thus wanted to know its spelling in order to pronounce it correctly in the oral summary.

One participant used fixation to think about the actual meaning of the word in relation to the context. She reported:

Quote 5.21

> I actually know what regression means, but in this context I think 'regression' means something else. It's a technical term, I guess. It's related to corruption and taxes. I stopped and tried to think what it means. I didn’t know; I couldn’t link this word to the (listening) content. (P1/*Corruption*)

In this case, the participant recognized the word ‘regression’ but thought that the meaning she knew did not go with the context, so she stopped and thought about alternative meanings that could fit the passage. It is worth noting that although fixation was used, it was not used successfully by these participants, either to identify words and spellings or meanings.

In addition to word recognition, fixation was reported to be used at the parsing stage to link words or series of words together to understand the meaning of utterances. The analysis showed that the participants were able to recognize single words, but because they did not understand these words in the context, they stopped and tried to make sense of their meaning in relation to the co-text (the immediately surrounding text). Participants, for example, stated:

Quote 5.22

> I was listening and taking note at the same time. I jotted down what I heard. At this point, I couldn’t follow it well. It's about 'muscle', ‘contraction’ and

> something related to ‘nerve transmission’. I stopped and tried to understand how these words are linked together and what they mean here. (P7/*Vitamin D*)

Quote 5.23

> I stopped here. It sounded like 'development system'. I was not sure. Then I heard 'ten times'. I was trying to think how these parts of information are connected to each other. (P4/*Corruption*)

In sum, the data show that fixation is generally used at two stages of processing, the perceptual and parsing stages. Nevertheless, fixation was used more for word recognition, to identify unfamiliar or unknown words, figure out word spellings, and generate accurate meanings of words in an unfamiliar context. At the parsing level, the participants used it to link individual words together to understand the meaning of utterances. The participants in this study did not seem to benefit from their use of fixation; however, as they still remained confused about the words and the meaning of the text they were trying to figure out.

5.2.3 Inferencing

Inferencing takes place when listeners use information within a text or conversational context to guess the meaning of unfamiliar language items, fill gaps in their listening, or link pieces of information together to build a more cohesive interpretation of the text. Inferencing relies on different sources of information such as, the already known, recognised words in an utterance, tone of voice and/background sounds, facial expressions, body language, or visual clues, and information beyond the local sentence level to guess the meaning of unknown words or parts of an utterance (see 2.4).

In the present study, the analyses show that inferencing was used by all participants across the four listening materials. It was reported to be used for three main purposes: 1) to understand the meaning of unknown or unfamiliar words, 2) to understand the meaning of sentences and single ideas of the passage, and 3) to infer links between pieces of information.

Two participants reported making inferences to understand the words that they were not familiar with. In listening to *Hans Krebs*, one participant did not know what ‘Warburg’ was. She then used the text coming after the word, which described an action of Warburg, i.e., ‘said the same thing’, to recognize that Warburg is a person’s name. This participant described:

Quote 5.24

> I was a bit lost when it talked about Warburg. I'm not sure if it was a person’s name or a place’s. I heard 'said the same thing'. I guessed Warburg is a person and he said something which I didn't know. (P11/*Hans Krebs*)

In listening to *Vitamin D*, one participant reported that she had not come across the term ‘tetany’ before, but in this listening she understood from the description ‘…and this results in a disease called tetany’ that it was a type of disease occurring when our body does not have enough Vitamin D. She said:

Quote 5.25

> I heard 'tetany'. I didn't know this word, but from this context I can say that it's a disease that occurs when we don't have enough Vitamin D. (P1/*Vitamin D*)

A major purpose of making inferences in the present study was to understand the meaning of utterances both at the sentence level and as a single concept by inferring relations between words and sentences. Some participants reported using visual cues together with the words or ideas in the text they could recognize, in order to arrive at the textual meaning that was unclear to them. They, for instance, indicated:

> *Quote 5.26*
>
> I saw two bars in the chart, one of which was about corruption and the corrupt payments and the other was about development assistance. I heard 'one trillion' and ‘the poor'. I was trying to link this text with the graph. I thought it was about the amount of money spent on corruption compared to that spent on the poor. (P1/*Corruption*)
>
> *Quote 5.27*
>
> I knew from the beginning of the listening passage that *Vitamin D* is related to blood circulation. I heard 'rapid die' or 'rapidly die', whatever.... I guessed that if we didn’t have enough Vitamin D, we would die rapidly. (P4/*Vitamin D*)
>
> *Quote 5.28*
>
> I heard 'workforce' and 'talent'. I inferred that talent is important to workforce. Then I heard 'modern economy', so I guessed that talent is also important to modern economy. I heard 'creative role' and I thought it must be about a creative role of talent in modern economy. (P4/*Talent*)

Some participants reported inferring the text they lost when they were concentrating on taking notes. They, for instance, reported:

> *Quote 5.29*
>
> I was taking notes here…. I missed the subject of the sentence….I wrote down 'overcome', 'human', and 'obstacle'. I knew that these words describe the man and he could be the subject of the sentence that I missed. (P5/*Hans Krebs*)
>
> *Quote 5.30*
>
> I was following the listening text and when it was about '9 milligrams per....', I noted the number…… I missed what followed after that. I guessed it was that if we had lower than 9 milligrams of Vitamin D, we would die. (P3/*Vitamin D*)

Some participants made inferences about the information they had missed while listening, for example, they said:

> *Quote 5.31*
>
> I missed it when the speaker talked about his father and his teacher, Warburg. I had no idea how these two people related to the story. Towards the end of the listening, it was 'children go on to do great things no matter what we say to them'. I thought Hans Krebs was inspired by these two people (his father and his teacher). (P1/*Hans Krebs*)
>
> *Quote 5.32*
>
> I heard that 'his father said' and then 'never be great scientist'. I later thought how these two parts of the text linked. I guessed it was what his father said to him, so in my summary I wrote that 'he always heard from his father and his lecture that he will never be a great scientist in Biochemistry'. (P9/*Hans Krebs*)

Another use of inferencing is to infer links between individual ideas to contextualize the information gained from the listening and create a background of the listening text for further processing. In *Talent*, the participants, for example, stated:

> *Quote 5.33*
>
> The listening passage was about talent in relation to its meaning. I heard 'anyone in top distribution' and 'anyone in the workforce'. I think the text is about the meaning of talent in business. (P10/*Talent*)
>
> *Quote 5.34*
>
> From what I heard such as define, talent, workforce, economy, I thought it was about definitions of talent. (P4/*Talent*)
>
> *Quote 5.35*
>
> It was about talent. I understood that it's the meaning of talent in business and talent in relation to business power. (P5/*Talent*)

In *Vitamin D*, one participant, for instance, indicated:

Quote 5.36

> I heard 'Vitamin D' and ‘the central function’. I inferred that it's about the advantages of Vitamin D. (P5/*Vitamin D*)

Some participants related words/sentences and ideas obtained from the listening to each other and situated the ideas in a wider context to understand the text’s overall theme. This occurred although some details had not been successfully understood. Examples are:

Quote 5.37

> I understood that the story emphasized a corruption problem. Development assistance is mentioned just to point out how much money is paid on corruption compared to development assistance. (P9/*Corruption*)

Quote 5.38

> In the beginning I heard 'talent' and then 'hard to define'. I told myself that it was about a definition of talent. (P2/*Talent*)

Quote 5.39

> I heard the word ‘talent’ and I saw a picture of human pyramid on the screen. I thought the text must be about human talent in an organization. (P3/*Talent*)

Some participants relied on the information they had gained from the listening to infer the main idea of the text. They, for instance, said:

Quote 5.40

> I noted down 'economic’, ‘social cost’, ‘one trillion dollars' and ‘affect to social cost’. I think the main idea is about the impact of corruption on social cost. (P2/*Corruption*)

Quote 5.41

> I didn't understand every word in the listening text, but I was trying to link what I had heard together. I got the point. I thought it was that corruption is a high cost and it has several impacts on the poor. (P8/*Corruption*)

Quote 5.42

> I knew that the main idea in this listening is that although people lack support from their parents, they can be successful. I am sure that it is this idea because the middle of the listening passage showed that 'he is a wonderful example of how human can overcome problems in life, and later there was some part of the text showing that his father and his teacher constantly discouraged him. (P3/*Hans Krebs*)

It is worth pointing out that, similar to fixation, inferencing was not always successfully used as can be seen in Quotation 5.31 in which the participant (P1) incorrectly inferred that Hans Krebs was inspired by his father and his teacher. In fact, the actual meaning of the text was that ‘Hans Krebs was discouraged by his father and his teacher, but he went on to do great things no matter what they said to him’. Another example of an incorrect inference is made by P1 in *Corruption* (see Quotation 5.26) as he inferred from the text and the image that the listening passage was comparing between corrupt payment and the poor. In fact, the listening passage compares the money spent on corruption worldwide with that spent on development aid. The attempts to infer links between pieces of information, however, appeared to be useful to some participants as they were able to quickly contextualize information they heard and used it for further text comprehension processing and to establish the main idea of the text. Inferencing thus seems to have been the crucial basis for successful listening in this study.

5.2.4 Elaboration

Elaboration is the strategy by which listeners use their prior knowledge from outside the text or conversational context and relate it to the text’s linguistic knowledge in order to compensate for missing information or embellish an interpretation to make the text more meaningful and complete (Goh, 2002 & Vandergrift, 1997). Sources of knowledge that are generally relied on in the application of this strategy include personal world knowledge, knowledge gained in academic situations, and a combination of questions and world knowledge to brainstorm logical possibilities.

The analysis showed that the highest number of the participants using this strategy (6 out of 12) was found for *Corruption*. Five participants used it when listening to *Hans Krebs* and two used it for their listening to *Talent* and *Vitamin D*. This strategy was applied mainly to understand the text above the sentence level. The types of knowledge the participants used to elaborate their understanding include their prior experience, academic background, and general knowledge. For example, some participants elaborated on the text that was not clear to them by relying on their prior experience. P4, for instance, reported:

Quote 5.43

> I couldn't catch all details in the listening. I used my background knowledge together with what I heard to understand it. This one was about corruption which has a high cost. I linked this idea to what happened in Thailand [the participant's home country], where corruption is a major problem. From my own experience, I can say that corruption has an impact on public services as a lot of money planned to spend on the public services goes to individuals' pockets. (P4/*Corruption*)

P8 also relied on his experience to elaborate on the text. He reported using elaboration both in his listening to *Corruption* and *Hans Krebs*. For *Corruption*, he indicated:

Quote 5.44

> My experience told me that corruption affects the poor more than the rich, so I inferred that when there is a high cost of corruption, the poor people will be directly affected. (P8/*Corruption*)

With reference to the *Hans Krebs* passage, he stated the following:

Quote 5.45

> I knew that the story was about a famous scientist. His name was Hans Krebs. Then it was about his father and his teacher. I generalized from my own experience that his father and his school teacher encouraged him to study hard and be a highly successful scientist. This is what my teacher and my dad did to me. (P8/*Hans Krebs*)

It is worth noting that in this case (Quote 5.45), elaboration was not used successfully because in the original text, Hans Krebs' father and his teacher did not encourage him; they did the opposite. This contrasted with the participant's own experience. Thus, the text elaboration in this act was not accurate.

In addition to prior experience, some participants relied on their academic background to elaborate on the text. P9, for instance, stated:

Quote 5.46

> I knew that corruption affects financial flows and this, of course, has an impact on the public. I study economy. It's quite easy for me to figure out what happens in a corruption cycle. (P9/*Corruption*)

In this case, elaboration was found to be useful as the participant appeared to be able to use his academic knowledge to elaborate on the text and understand it correctly. This is because the participant's knowledge is in line with the text message.

Elaboration was also made from the participants' general knowledge, as can be seen in the following quotes:

Quote 5.47

> I depended a lot on my background knowledge to create a summary as I got a feeling that the audio text and the graph didn't go together. Here I summarized that corruption impact the poor more than the rich. I think it is a general fact that we all know. (P7/*Corruption*)

Quote 5.48

> I knew it was about a high cost of corruption which is about 10 times higher than assistant development. I then thought based on my knowledge that corruption directly affects the poor and creates inequality in the society. (P2/*Corruption*)

In *Corruption,* elaboration which relied on the participants' general knowledge was done successfully as the participants understood that *Corruption* had an impact on the poor more than the rich, which was true to the text.

In *Talent* and *Vitamin D*, elaboration was also beneficial to the participants as the elaborated information appeared to be related to the topic theme. P4, for instance, explained in his stimulated recall that he thought talent in the modern world is associated with being creative, which is partly true according to the text. In his own words, he said:

Quote 5.49

> I heard 'creative' and 'modern economy'. I thought these two things linked to each other as you can see in the modern world, people with creativity are very important to an organization and businesses. I provided in my summary that the speaker emphasized the importance of a creative role of talent in modern economy. (P4/*Talent*)

In *Vitamin D,* where the theme of the topic is about the benefits of *Vitamin D* and problems if people do not have enough of it, elaboration was also found useful. It assisted the participants to understand the main point of the listening passage better. The participants reported:

Quote 5.50

> I understood that the listening passage was about the benefits of Vitamin D and its main function is to maintain blood calcium. I know that when people don't have enough calcium, they will have a problem with their muscle and bones, so I put in my summary that having calcium lower than 9 milligrams per 100 milliliter blood can affect muscles. (P9/*Vitamin D*)

Quote 5.51

> When I listened to Vitamin D, I was trying to think how in general people talk about Vitamin D. In the same way as it was in this listening text, they are likely to discuss about the benefits of Vitamin D. (P5/*Vitamin D*)

Unlike in *Corruption*, *Talent*, and *Vitamin D* where elaboration tended to be applied successfully, in *Hans Krebs* elaboration was found to negatively affect the participants' understanding of the text. The participants' experience was that parents and teachers support their children to do great things in life. In contrast, the main theme of *Hans Krebs* is that children can be successful even though they lack support from parents or teachers. Thus, the activation of elaboration on the irrelevant knowledge/experience was not useful in this listening topic. This can be seen in the following examples:

Quote 5.52

> It was about a great scientist, children, and a great achievement. Based on these words I thought that he (Hans Krebs) inspired kids to study science and do great things in their life. This is because great or the story of famous people is in general used to inspire children to study and do things to be successful in life. (P7/*Hans Krebs*)

Quote 5.53

> The listening text was about a great scientist, his family background, and his school life. As we know that family plays an important part in individual's success. In this case, I think it's the same, so I concluded in this story that his success is devoted to the family. (P7/*Hans Krebs*)

In sum, the use of elaboration in this study appeared to be based on several types of knowledge, including personal experience, academic background, and general knowledge of the world. Success in elaboration, however, depended to a large extent on the relevance of the knowledge or background the participants brought to the tasks. When listening to *Hans Krebs*, in particular, elaboration was disadvantageous since the participants' experience and background were not in line with the listening theme.

5.2.5 Reconstruction

Reconstruction is conducted when listeners miss some part of a text, hear only part of it and rely on words/phrases gained from the text to recreate it and make their comprehension complete. The outcome is usually the gist of the text they have just heard, which could be either relevant or irrelevant to the original message, depending on whether the words/ideas they have captured are part of the text's necessary information (Goh, 2002). This strategy differs from inferencing in that inferencing occurs during listening while the listeners are building up their understanding of the text. Reconstruction, on the other hand, is used towards the end of the listening when the listeners feel that they still have not got the main point of the listening but have to recreate the text to complete the tasks.

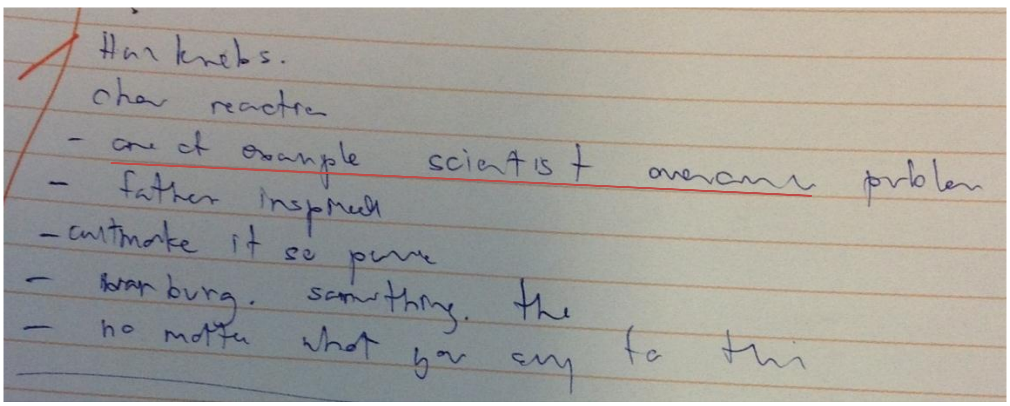
In this study, reconstruction was reported only in the listening to *Hans Krebs* and *Corruption*. In *Hans Krebs*, where the highest number of the participants using reconstruction was found (9 out of 12), the participants reported that they had missed some parts of the text or had problems understanding the main points despite the fact that they had recognized certain words/phrases. As a result, they ended up relying on those words to establish the main idea or part of the text that they had missed to complete their summary. P10, for instance, said:

Quote 5.54

> I didn't really get the points in the listening. I am not familiar with this name (Hans Krebs) and I had no idea who this man was. I remember that I heard 'no matter what we say to them blah, blah, blah.... I think this is the main point, but because I missed the previous part of the text, I had to make up a concept from the words I had to replace what I had missed. (P10/*Hans Krebs*)

The analysis of this participant's notes and summary content also showed that she relied on her notes to fill in missing information. In her notes, she wrote:

Note 5.1



(P10/*Hans Krebs*)

In her oral summary, this participant combined the notes 'one of example scientist overcome problem' and 'father inspired' to recreate the message that 'he was inspired by his father'. She orally summarized it as follows:

Oral Summary 5.1

> …his ability (Hans Krebs' ability) was inspired by his father umm father motivation that don't let anyone like say that you cannot do anything or something is impossible to do it and when he went to school at Walberg, the teacher also talking about. (P10/*Hans Krebs*)

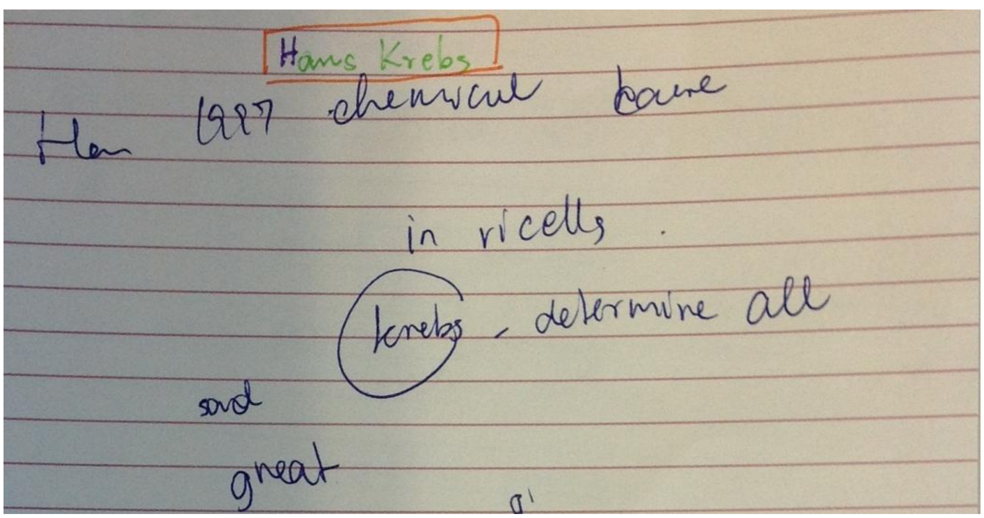
Some participants admitted that they did not understand the point the speaker was trying to make in the spoken text, they jotted down the words/phrases and relied on those notes to create a story. P4, for instance, reported:

Quote 5.55

> I didn't really understand the passage so what I mainly did is to try to catch as many words as possible and then I relied on these words to make a story. (P4/*Hans Krebs*)

This participant's notes seem to show that he did not catch much of this listening. He did not seem to understand the point the speaker was trying to make. His notes contain a few words only:

Note 5.2



(P4/*Hans Krebs*)

His oral summary also shows that the participant reconstructed the message in order to complete the task. However, as several key words and points had been missed, his summary (5.2) did not successfully reflect the content of the original text. The summary failed to include the text's main idea which is 'Hans Krebs is a wonderful example of people who can overcome all kinds of human obstacles'.

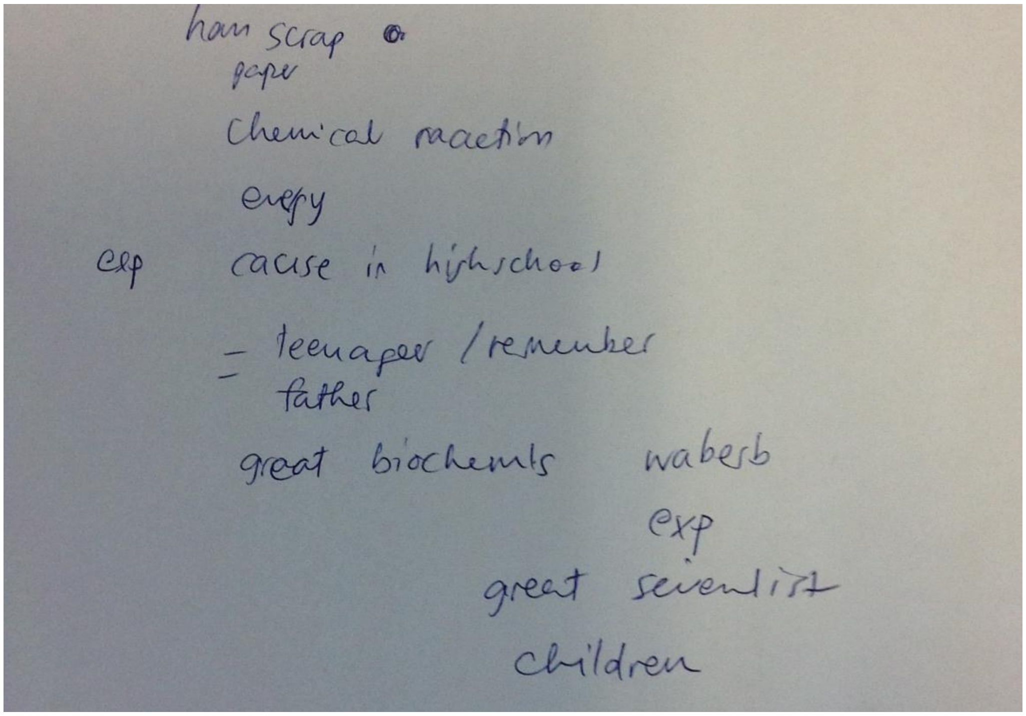
> *Oral Summary 5.2*
>
> Um.. Krebs is one of the greatest um.. chemist.. chemistry researcher, I think and I think he is ..um and the writer talked about um his great success and the influence ...his influence on children and now I think he is maybe and inspiration ..has a great inspiration on children. (P4/*Hans Krebs*)

Another participant who failed to provide an accurate summary of the listening text when applying the reconstruction strategy is P6. This participant reported in her stimulated recall.

> *Quote 5.56*
>
> I took notes of all key words. Then I created a story from these words. (P6/*Hans Krebs*)

In her notes, she wrote:

Note 5.3



(P6/*Hans Krebs*)

Although the words/phrases written down were stated in the listening passage, they did not appear to express the key idea of the text, namely, that children can be successful without parental support. In addition, no links between the words seemed to be made by the participant. As a result, the oral summary provided by this participant was not accurate, as can be seen in oral summary as follows.

Oral Summary 5.3

I heard a lecture about Hans Krebs, who gave example err on a paper about chemical reaction. He gave example about cause and effect when he studies in high school and when he was teenager and with a memory about his father and he also umm gave example about great bio-chemical and great scientist when... an experience of great scientist when they were young. (P6/*Hans Krebs*)

Another participant who reported using reconstruction when listening to *Hans Krebs* is P12. The main reason for this participant having to reconstruct the original message, as indicated in his stimulated recall, is because he could not link pieces of parsed information together. Therefore, he recreated a summary of the spoken text by using the words/parsed information he had gained. He indicated:

> *Quote 5.57*
>
> I was very confused in this listening (*Hans Krebs*). I didn't know how it relates to a chemistry course. …I couldn't link the information together. I heard ‘career’, ‘his father told him’, ‘overcome human obstacle’. I had to make up the story from these words to complete the task. (P12/*Hans Krebs*)

In his oral summary, he stated:

> *Oral Summary 5.4*
>
> Yes, this article is talking about Hans Krebs that he has studied before and he tried to ah show how to overcome human obstacle like ability or anything that we have before and ah he showed example of children do something and they had some stacles and ah… (P12/*Hans Krebs*)

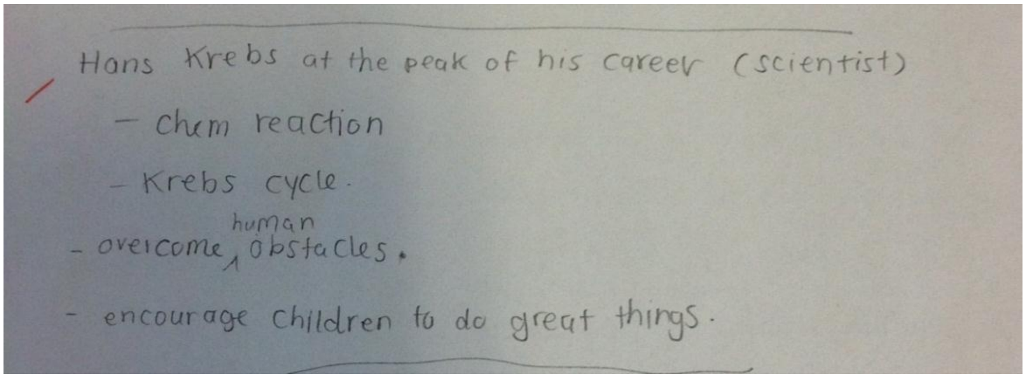
In listening to *Hans Krebs,* one participant reported being lost while listening and missing the first part of the spoken text. However, she managed to get back to the listening text, just in time to hear ‘he is the wonderful example of how people can overcome all kinds of obstacles’, which is a key idea in the text, and managed to include it in her summary. She reported:

> *Quote 5.58*
>
> I thought in the beginning of the listening, the story was about the life of this man (Hans Krebs). …I lost. I couldn’t understand what I was listening to. It was blank in my head. I heard something about his work, Krebs' cycle. …I was lost again. ….now I could get back to the text ... The speaker said 'he is the wonderful

example of how people can overcome all kinds of obstacles'. I was really confused how this part "all kinds of obstacles' came. (P11/*Hans Krebs*)

In her notes, she wrote:

Note 5.4



Hans Krebs at the peak of his career (scientist)

- chem reaction
- Krebs cycle.
- overcome human obstacles.
- encourage children to do great things.

(P11/*Hans Krebs*)

As the participant admitted she had missed part of the text, she relied on the words/phrases she could catch to replace the missing part. However, although missing several points, this participant could recognize one key point of the text which is 'overcome human obstacles'. Consequently, she was to some degree successful in reconstructing the original message. She was able to state a key idea in her summary correctly, namely that 'he can overcome all the obstacles in his life and encourage people to do great thing'. Her oral summary was:

Oral Summary 5.5

Umm, this guy said that he has Hans Krebs, the scientist who came up with the chemical reaction theory, to be his role model in terms of how he can overcome all the obstacle in his life and encourage people to do great thing. You can do everything. (P11/*Hans Krebs*)

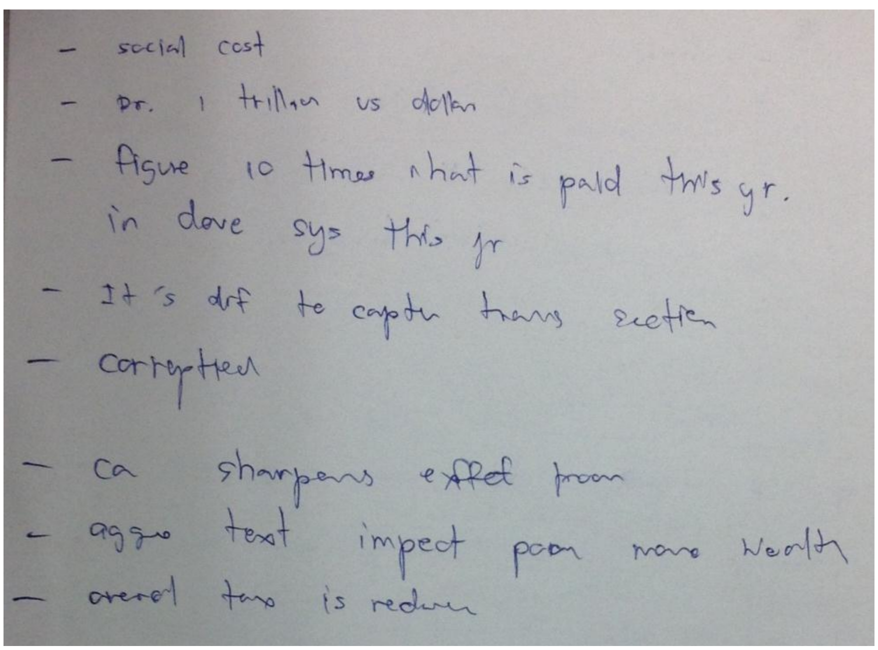
With reference to *Corruption*, the participants also used reconstruction to recreate part of the text they had missed and to derive the main point of the listening. P10, for instance, stated that she did not really understand what the passage was about. She used the words that she could detect from the text to reconstruct her summary. In the stimulated recall, she indicated:

Quote 5.59

> I didn't really understand what the passage was about. I heard 'social cost' but had no idea what the point was the speaker wanted to make. (P10/*Corruption*)

In her notes, she wrote down the words/sentences she heard:

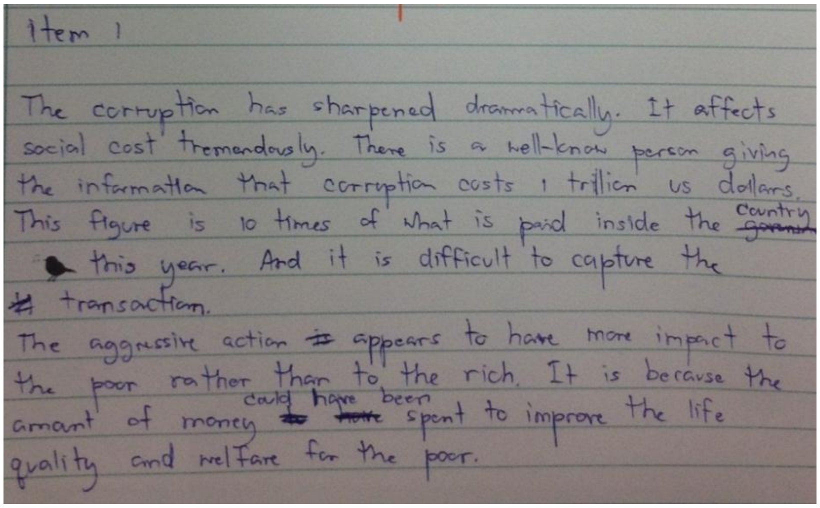
Note 5.5



(P10/*Corruption*)

By combining these words, she wrote the following summary:

Written Summary 5.1



Item 1

The corruption has sharpened dramatically. It affects social cost tremendously. There is a well-know person giving the information that corruption costs 1 trillion US dollars. This figure is 10 times of what is paid inside the ~~govern~~ country this year. And it is difficult to capture the ~~#~~ transaction.

The aggressive action ~~is~~ appears to have more impact to the poor rather than to the rich. It is because the amount of money ~~to~~ ~~have~~ could have been spent to improve the life quality and welfare for the poor.

(P10/*Corruption*)

From the words/sentences she had captured, this participant was not successful in reconstructing the story as her summary did not convey key ideas of the text such as ‘corruption has impact on social and economic cost’.

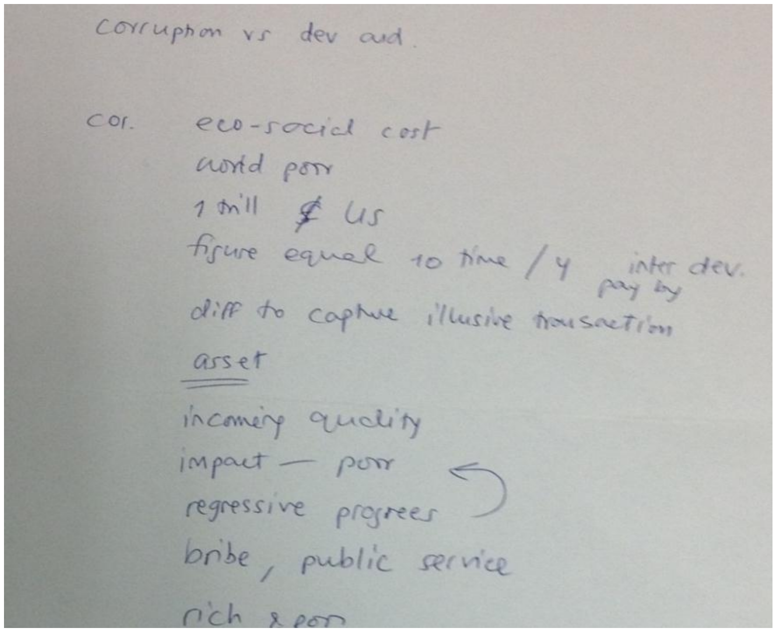
P6 is another participant who had difficulty understanding *Corruption* and thus relied on the words/sentences she wrote in her notes to reconstruct the story. She described it in her stimulated recall as follows:

Quote 5.60

This topic (*Corruption*) is hard. I knew it was about corruption, public services, the poor did not have access to public service, a number one trillions US dollars, ten times, capture transaction etc. However, I didn't catch the main point. I was a bit frustrated. I couldn’t differentiate which were facts and which were the speaker's opinions. I blended everything together and wrote a summary. (P6/*Corruption*)

In her notes she wrote:

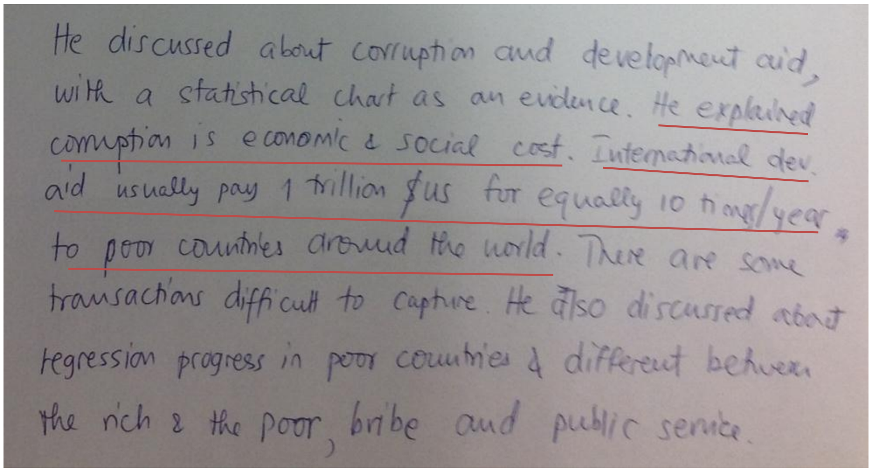
Note 5.6



(P6/*Corruption*)

Although this participant stated a general idea in the second sentence of her summary correctly (Written summary 5.2; ‘corruption is economic and social cost’), she provided incorrect supporting details in the later part of her summary. She wrote that ‘international development aid usually pay 1 trillion $US for equally 10 times/year to poor countries around the world’. However, this information does not match the original text which describes that ‘about one trillion $US is spent on corruption each year and this amount is ten times higher than development assistance’.

Written Summary 5.2


He discussed about corruption and development aid, with a statistical chart as an evidence. He explained corruption is economic & social cost. International dev. aid usually pay 1 trillion $us for equally 10 times/year, to poor countries around the world. There are some transactions difficult to capture. He also discussed about regression progress in poor countries & different between the rich & the poor, bribe and public service.

(P6/*Corruption*)

To summarize, the analysis showed that the reconstructions made by the participants mainly followed a bottom-up pattern. They were based on words and phrases that the participants had captured from the text, but most of these words/phrases did not convey the text's main point. Consequently, the reconstructed information was likely to be unproductive and misrepresent the key point of the listening text. It is worth noting that in *Talent* and *Vitamin D*, there was no record of the participants using the reconstruction strategy. This might be because the participants were able to contextualize and understand most of the ideas in these two listening passages.

5.3 Overall picture of the cognitive strategies used at each level of cognitive processing

Table 5.1 maps the five cognitive strategies used by the participants against the cognitive processes which they assist and against the different levels of listening processing for which they aim to solve comprehension problems.

Comprehension processing stages	**Cognitive processes**	**Strategic processing**
		Cognitive strategies
Lower-level processes (linguistic processing)	Acoustic-phonetic processing	Fixation
	Word recognition	Fixation Inferencing Reconstruction
	Parsing	
Higher-level processes (comprehension processing)	Semantic processing (at the local level) Semantic processing (at the global level) Pragmatic processing	Prediction* Inferencing Elaboration Reconstruction

Table 5.1: Overall picture of cognitive strategies used at each level of cognitive processing

**Although used at the higher-level, prediction was for facilitating listening processing by predicting the upcoming text, not for solving listening comprehension problems*

The cognitive strategies evidenced in the dataset are prediction, fixation, inferencing, elaboration, and reconstruction. As can be seen in Table 5.1, three of these - fixation, inferencing, and reconstruction - were employed both at the lower- and the higher- levels of processing. Prediction and elaboration were however found at the higher-level processes only.

Two of the cognitive strategy types, i.e., fixation and reconstruction, were used mainly in a bottom-up direction. Fixation is where the participants stopped to focus on sounds to recognize words, spellings, or meanings. As one of the least used strategies, it might suggest that in general the participants did not have much difficulty processing the texts' lexicon. Or, it could be that the participants ignored the problematic words and directed their attention away from these and focused on other elements of the text which they considered more important. Reconstruction is another strategy used by a small number of the participants and also used in a bottom-up manner. It took place when the participants relied on the words/sentences to reconstruct the original message. In *Corruption* and *Hans Krebs*, about half and more than half of the participants, respectively, reconstructed the text based on the words/sentences they could detect. This is partly because they were not able to follow the texts properly and missed most of the key ideas in the listening passages, ending up reconstructing the texts to replace those ideas and fulfill the task requirement of providing a summary either in oral or written form. The use of this strategy did not appear to be useful to the participants as the content they reconstructed mostly inaccurately reflected the original texts.

The use of the inferencing, elaboration, and prediction strategies appeared likely to follow a top-down process pattern. That is, instead of relying heavily on the text's linguistic information, the participants brought in their prior experience, topical knowledge, and knowledge of the world to combine with their partial understanding of the text. However, these strategies were only useful if the participant's background knowledge and prior experience corresponded with the texts' information. Relying heavily on background knowledge and experience in listening to the topic where one's knowledge and experience deviated from the text's information, and not paying enough attention to the text's linguistic information proved to be disadvantageous, as was the case for *Hans Krebs*.

It is interesting to note in the case of *Talent* and *Vitamin D* that no participants employed the reconstruction strategy. This might be because everyone was able to make effective inferences and contextualize the text they were listening to, and process it for comprehension effectively. In *Hans Krebs* and *Corruption*, on the other hand, about half of the participants used reconstruction strategies. This might be because they were not successfully processing the texts for meaning, and thus, in order to complete the tasks, they had to reconstruct the text's meaning based on lexical and parsed information they managed to capture.

5.4 Summary

This chapter presented the cognitive strategies used by 12 participants to solve problems in listening. A total of five cognitive strategies were reported being used. Presented according to their occurrence from the beginning of the listening tasks to the end these are: 1) prediction, 2) fixation, 3) inferencing, 4) elaboration, and 5) reconstruction. Comparing the cognitive strategies used, inferencing was employed by the highest number of participants, as in fact, every participant reported using it while listening to every listening input. Reconstruction was the least used strategy. Fixation was used mainly at lower-level processes and the others were used both at the lower- and the higher-level of cognitive processing. Although these strategies were used by the participants, this was not always done successfully. Success in the use of the strategies at the higher-level (prediction, inferencing, elaboration, and reconstruction) depended partly on the participants' ability to process the text at the lower-level.

Chapter 6 Metacognitive strategies

6.1 Introduction

The use of metacognitive strategies is another important aspect contributing to the ability to communicate in a second language, in addition to cognitive processes and cognitive strategies (Bachman & Palmer, 2010; Vandergrift, 2003). Although metacognitive strategies are not used as directly to understand texts as cognitive strategies are, they provide a management function for text processing (Goh, 2002). That is, they are used to oversee the listening process, assist users in activating appropriate cognitive processes, and cope with difficulties arising from the text or other related aspects of task performance (e.g., nervousness). Specifically in listening tasks, the use of metacognitive strategies has been found to involve several aspects of task processing management, including 1) analyzing the requirements of tasks and setting goals, 2) activating appropriate cognitive processes for listening, 3) verifying predictions when they are incongruent with the text, 4) monitoring listening comprehension, and 5) evaluating the successfulness of the listening approach (Goh, 2002; Vandergrift, 2003). Under metacognitive management, mental or cognitive processes are expected to proceed more efficiently.

This chapter provides the findings on the metacognitive strategies the participants used to manage their listening process in the listening-to-summarize tasks. As discussed in Chapter 2, similar to cognitive strategies, metacognitive strategies are thought to be employed with some degree of consciousness (Vandergrift & Goh, 2012). To gather information on the use of metacognitive strategies, the stimulated recalls were analyzed, supplemented by the participants' notes and summary contents where relevant.

6.2 Metacognitive Strategies

Relying on previous research on L2 listening comprehension (Goh, 2002; O'Malley et al., 1989; Vandergrift, 2003; Vandergrift & Goh, 2012), metacognitive strategies in this study are defined as the conscious processes that the participants used to oversee and manage cognitive processes and strategies (see 2.4.3). The analysis showed that a total of six metacognitive strategies were used by the participants to monitor their listening process in the listening-to-summarize tasks (see Figure 6.1). The strategy presented on the far left of the X-axis in Figure 6.1 is 'preparing for listening', which the participants applied at the initial stage of listening task performance, i.e. after the tasks were introduced and before the audio file started. The strategies activated while the participants were listening to the audio, were 'selective attention', 'directed attention', 'comprehension-monitoring', and 'real-time assessment of input strategies'. On the far right of the X-axis is 'comprehension evaluation' which was employed towards the end of the listening.

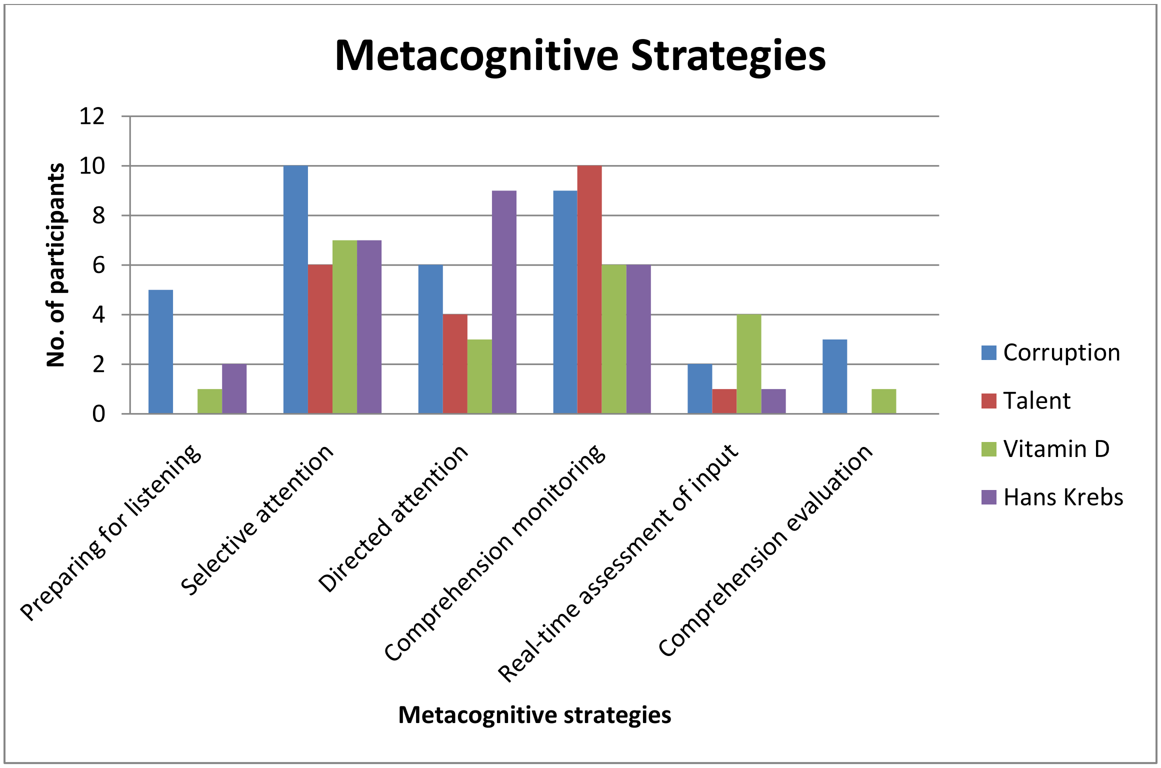


Figure 6.1: Metacognitive strategies used by the participants to monitor listening comprehension processing

As shown in the figure, the use of metacognitive strategies appeared to vary in terms of number of participants and according to the listening topics. Comprehension monitoring was used by the highest number of participants, followed by selective attention and directed attention, respectively. Less than half of the participants reported applying comprehension evaluation, real-time assessment of input, and preparing for listening strategies. A detailed description of each metacognitive strategy use is presented in the following sections.

6.2.1 Preparing for listening

Preparing for listening, as it has been named, occurs at the beginning of a listening task (Vandergrift & Goh, 2012). This is when listeners prepare themselves mentally and emotionally for the listening tasks. Goh (2002) and Vandergrift (2003) found that preparing for listening is activated for different purposes, including determining the requirements of the listening task, dealing with anxiety, anticipating key or content words, and rehearsing sounds of potential key words. In this study, almost half of the participants prepared to listen in the *Corruption* task (5 out of 12). Two participants applied this strategy at the start of the *Hans Krebs* task and one participant at the start of the *Vitamin D* task. No participant reported using it for *Talent*.

P6, who reported applying pre-listening preparation in the *Vitamin D* task, also did it for the *Hans Krebs* and the *Corruption* tasks. With respect to *Vitamin D*, she rehearsed the task instructions and anticipated what the passage would be about by relying on visual information on the screen. She explained:

> *Quote 6.1*
>
> I was waiting for the audio to play. I reminded myself that I was going to listen to a lecture….I saw a picture of Vitamin D. I thought it was about the value of Vitamin D. (P6/*Vitamin D*)

While listening to *Hans Krebs*, this participant was trying to eliminate her anxiety and re-read the task instructions. She stated:

> *Quote 6.2*
>
> I was trying to calm myself down and getting ready for the listening. I couldn't concentrate, ….[When the speaker said Item 1] I looked at the screen and read the instructions again. (P6/*Hans Krebs*)

In *Corruption*, where she complained that the instructions were long, this participant reminded herself what she had to do to complete the task:

Quote 6.3

> [Before the audio started] I sequenced in my head what I was supposed to do, first I listen to a lecture, and after that write a summary of at least 50 words to summary the listening in 10 minutes. (P6/*Corruption*)

What this participant did before listening was mainly reminding herself of the task instructions, setting goals for her listening, and diminishing her nervousness. However, she did not report applying this strategy in her listening to *Talent*. This might be because *Talent* was the last task delivered to her, so she might not have been nervous any longer nor unclear about what she was supposed to do to complete the task. This might suggest that the use of this strategy may not be task-dependent, but practice-dependent. Once test-takers know what they are supposed to do in that particular task, they may no longer be nervous about the task or spend time preparing for listening.

Another participant who reported preparing before listening is P5. She used it to make herself mentally ready for the *Hans Krebs* text. She reported:

Quote 6.4

> [Before listening] I was trying to concentrate, trying not to be nervous. [When a picture was presented on the screen] I looked at the picture. I got a name, 'Hans Krebs'. (P5/*Hans Krebs*)

In *Corruption*, pre-listening preparation was performed for three purposes: to 1) rehearse the task instructions, 2) lower anxiety, and 3) evaluate the participant's own background knowledge of the topic. The participants indicated:

Quote 6.5

> I was trying to concentrate on the task. I reminded myself what I had to do…I saw a graph. I self-evaluated my background knowledge of the topic. It seemed to be about social science, not pure science. I felt a bit relieved. (P2/*Corruption*)

Quote 6.6

> I looked at the graph and thought how much knowledge I had of the topic. I thought it would not be difficult for me. I felt relaxed. (P8/*Corruption*)

One participant directed his attention to the listening and waited for the audio file to start. He said:

Quote 6.7

> Here, I looked at the picture but I couldn't understand it. I thought I needed an explanation. I was waiting for the audio to play…..concentrating and getting ready to listen. (P3/*Corruption*)

To summarize, preparing for listening in this study was performed with three main purposes: to 1) remind oneself about the task instructions and set a listening goal, 2) eliminate anxiety, and 3) evaluate one's own topical knowledge. Most of the participants reported reminding themselves of the task instructions, especially in the first task they did. The number of participants who engaged in this strategy, however, is far below that of those who did not use it. This might be because the instructions were delivered to the participants step-by-step and two sample items were given prior to the actual tasks. Thus, the participants may have been aware of what they had to do on the basis of the sample tasks. However, although some participants had understood the task instructions without too much trouble and remembered what they had to do to complete the tasks, they wanted to re-evaluate their own understanding and ensure that they set the right goals for task completion. Furthermore, a relatively small number of participants indicated that they were trying to diminish their anxiety before listening which seemed related to knowing what they had to do (the instructions) and feeling unsure about whether they would understand the listening content. Some participants reported that they were very worried about their ability to understand the listening tasks and got so nervous that they had to calm themselves down. Some evaluated their topical knowledge to see whether it would assist them to successfully perform the tasks. Since the participants used this strategy to get ready for the listening, preparing for the listening can be regarded as a useful metacognitive strategy.

6.2.2 Selective attention

Selective attention occurs when the listeners notice or pay attention selectively to specific aspects of listening input. According to Goh (2002), listeners use this strategy at least for two reasons: to 1) catch parts of the text which are considered important and contribute to understanding of the topic and 2) direct attention rapidly to the main point, without having to pay attention to all details.

In this study, the highest number of participants applying selective attention was found for *Corruption* (10 out of 12). The second highest number was found for *Vitamin D* and *Hans Krebs* (seven participants using it in each listening). Six participants applied this strategy when listening to *Talent*. Selective attention was used in combination with the cognitive strategy of prediction. That is, once the participants saw the picture related to the listening text on the screen and started listening to a couple of sentences, they anticipated the theme of the story they were going to hear. During this process, it is likely that their schemata related to the topic were activated to process the incoming text. Their attention was then selectively directed towards those text elements that related to what they had anticipated. For example, in the *Corruption* text, where a bar chart was provided, the participants predicted that the text was going to compare two things. While listening, they paid attention to the two bars and listened for words or numbers that they thought would be used for describing the chart. Below are excerpts from participants' verbal reports.

> *Quote 6.8*
>
> I looked at the graph. I tried to detect the words in the listening that describe it. I got 'economic', 'social cost', and 'one trillion dollars'. I heard 'ten times'…I was thinking how it was related to the graph and I wanted to know how it was related to the social cost. (P2/*Corruption*)
>
> *Quote 6.9*
>
> I focused on the graph. It [the graph] contained two bars. One was higher than the other. This showed that the speaker was making a comparison between two things. I heard ten times. I noted it down. It was a number. I knew that I needed it to describe the graph. (P11/*Corruption*)
>
> *Quote 6.10*
>
> When the graph was shown, I focused on it. It was about 'financial' and 'corruption'. When the audio started, I noted down words….I searched for other vocabulary that could link to the graph…I intentionally focused on figure. I heard 'ten times' and 'one trillion dollars'. I noted them down. (P12/*Corruption*)

Some participants predicted that the speaker was going to compare the two bars by using different numbers. Therefore, they looked out for numbers that the speaker would mention in the text to describe the chart. They, for example, said:

> *Quote 6.11*
>
> I was trying to catch numbers used to describe the graph. I got 'one trillion', ten times',.... (P1/*Corruption*)
>
> *Quote 6.12*
>
> I was identifying the context of the listening, trying to catch the ideas describing the context. It was about 'bribery' and 'the amount of one trillion dollars'. I noted them down because I thought they were associated with the graph. (P3/*Corruption*)

Quote 6.13

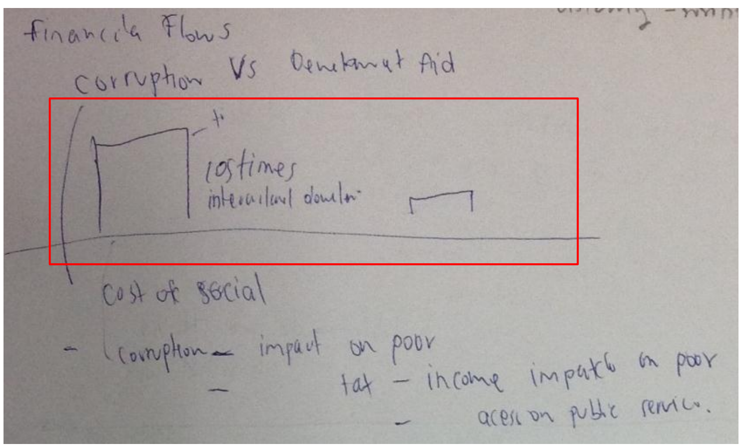
> I focused on numbers. I was thinking that I should have heard more numbers because it was a graph which was usually described by numbers. (P7/*Corruption*)

Quote 6.14

> I was trying to catch how much money paid on corruption and on development assistance. (P8/*Corruption*)

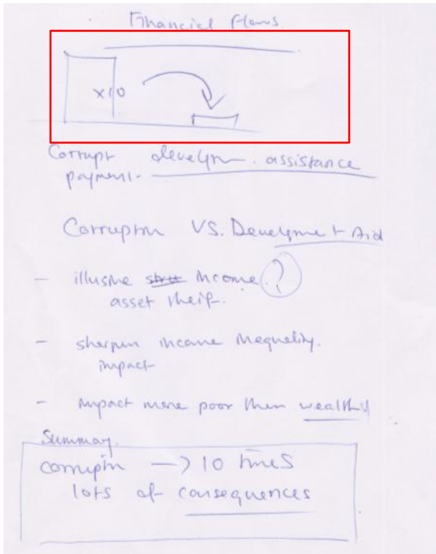
The following notes also demonstrate that the participants concentrated on the graphic information and numbers as they drew graphs in their notes and noted down some numbers they heard.

Note 6.1



(P5/*Corruption*)

Note 6.2



(P8/*Corruption*)

While listening to *Hans Krebs*, where a picture of a man was presented on the screen, the participants focused on the picture and the description below it. Some participants predicted that they were going to find out more about the man's story. As a result, they selectively paid attention to words used to describe people, i.e. names, work, and success. Some of the participants reported:

Quote 6.15

> I focused on his name and what he did before. I was paying attention to his career which I thought was the main point of the listening passage. (P5/*Hans Krebs*)

Quote 6.16

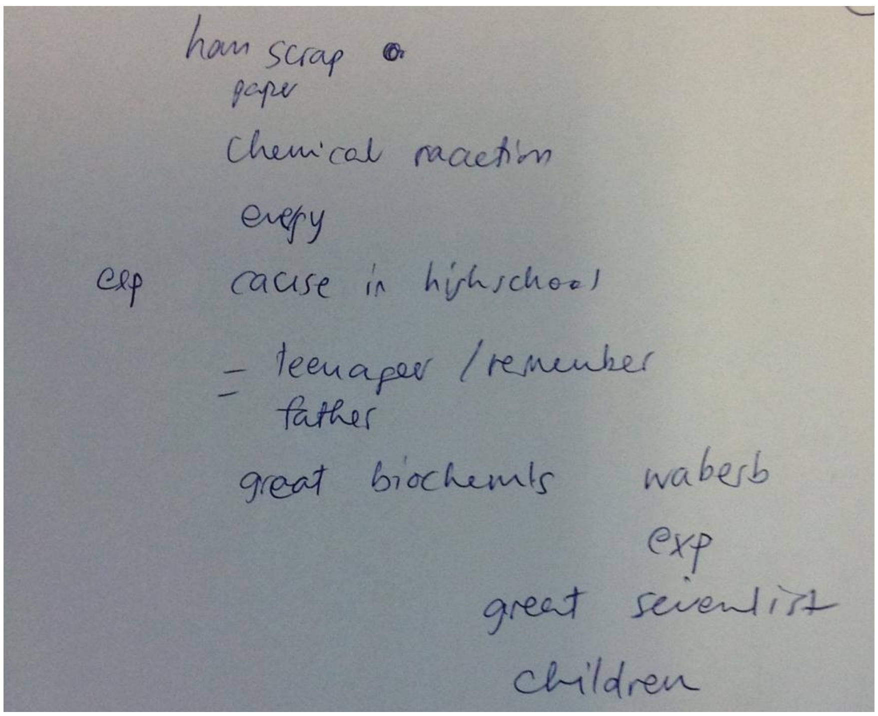
> I focused on the picture,… his name, and made sure that I got his name correct. I kept on listening and listened for vocabulary relating to his biography and his career. (P2/*Hans Krebs*)

Quote 6.17

> I looked at the picture. I got his name, Hans Krebs. I thought the listening passage was about this man and his career…. I focused on vocabulary used to describe his job. (P8/*Hans Krebs*)

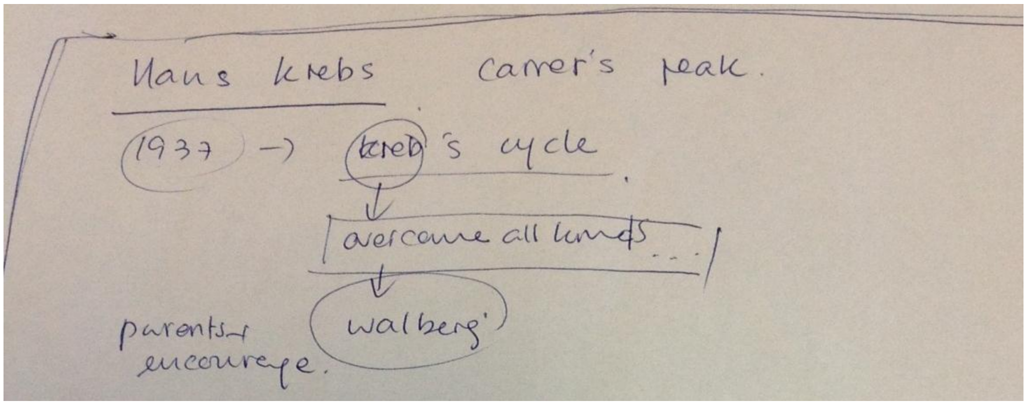
The notes also support that the participants attached importance to personal information mentioned in *Hans Krebs*, i.e. his name and his work. For example:

Note 6.3



(P3/*Hans Krebs*)

Note 6.4



(P8/*Hans Krebs*)

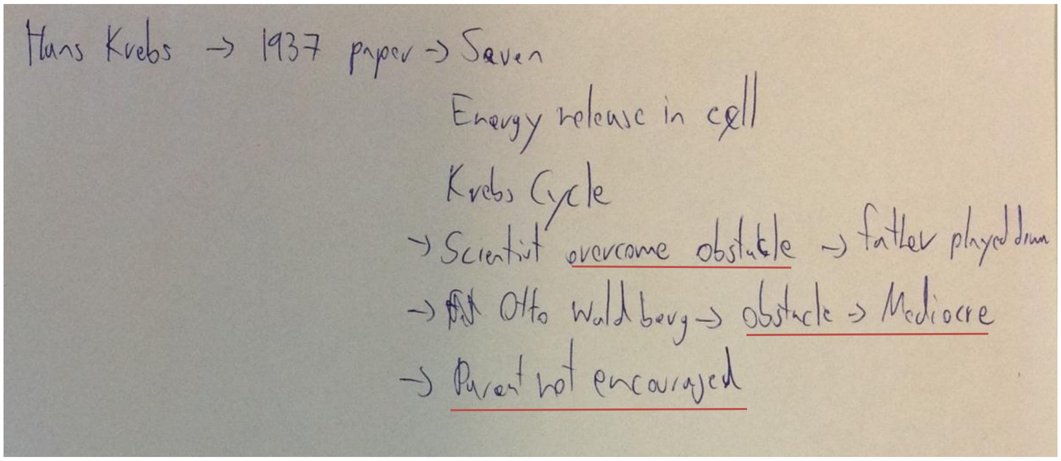
One participant thought that the main point for *Hans Krebs* was not Krebs' biography but the way he overcame a number of obstacles in his career development. The participant therefore kept his focus on this direction instead of on his biography, as shown in the following quote.

Quote 6.18

When it came to 'overcome obstacle', I knew that this was the focus of the story. I then focused on the vocabulary used to describe how he overcame his obstacles. I noted down key words such as 'mediocre scientist' and 'not encouraged'. (P3/*Hans Krebs*)

In his notes the participant wrote:

Note 6.5



(P3/*Hans Krebs*)

While listening to *Talent* and hearing that ‘there are different ways to define talent’, the participants reported realizing that the text was mainly about the meaning of talent. Thus, they specifically focused on vocabulary used to give definitions such as ‘is’, ‘mean’, and ‘define’. They, for example, said:

Quote 6.19

I noted down ‘talent’, ‘difficult to define’, and ‘different views'. (P2/*Talent*)

Quote 6.20

When I heard 'what is talent?’, I noted it down. I noted down 'different', ‘difficult', ‘anyone in top distribution’, ‘anyone in the workforce top restrictive’. I think these words were important to define talent. (P10/*Talent*)

Quote 6.21

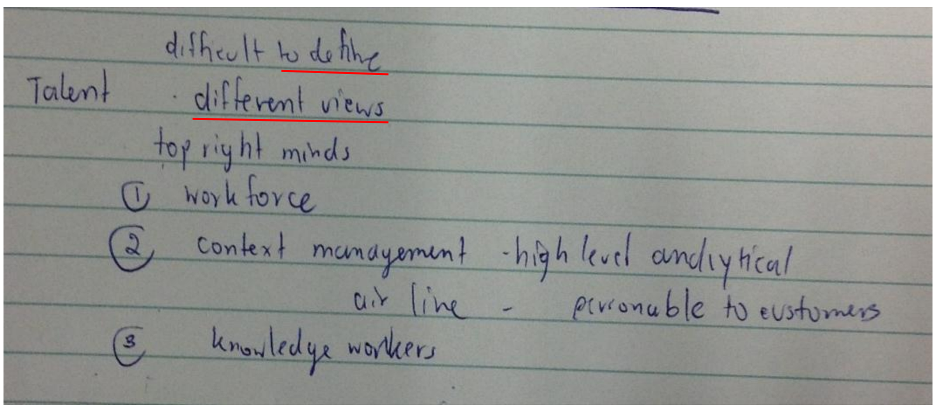
When I heard 'definition', I knew that it was about giving definitions. I focused on words, such as ‘define’, ‘definition’, ‘terms’..… etc. (P11/*Talent*)

Quote 6.22

> It was about a definition. I paid attention to the words that were used to define terms, such as 'is' or ‘definitions’. They helped me to get directly to the main point. (P7/*Talent*)

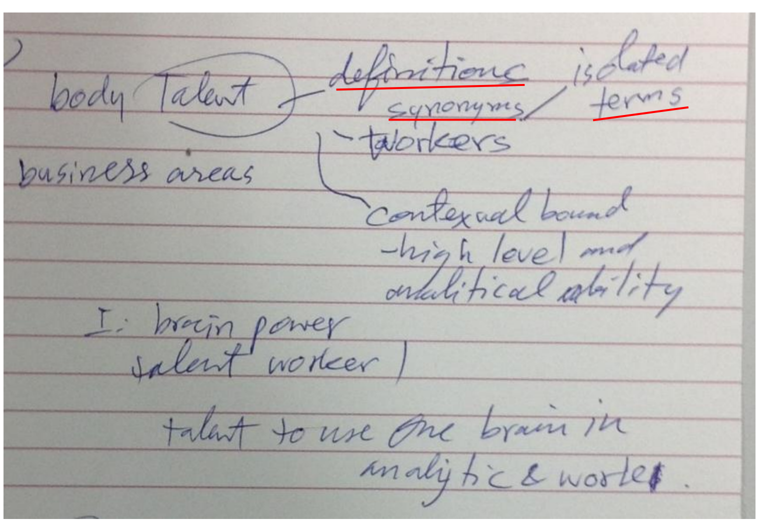
The notes also support that the participants paid particular attention to the words used to define the terms. For example:

Note 6.6



(P2/*Talent*)

Note 6.7



(P7/*Talent*)

In *Vitamin D,* where the participants understood that the main point was about the benefits of Vitamin D, the participants paid attention to the vocabulary generally used to describe health and health problems and other terms used in health science.

Quote 6.23

> I focused on the words related to health and health problems. I got ‘blood calcium level’ and ‘muscle nerve’. (P6/*Vitamin D*)

Some participants also felt that *Vitamin D* sounded more scientific than the other listening texts. Therefore, they focused heavily on the technical terms used in science while listening.

Quote 6.24

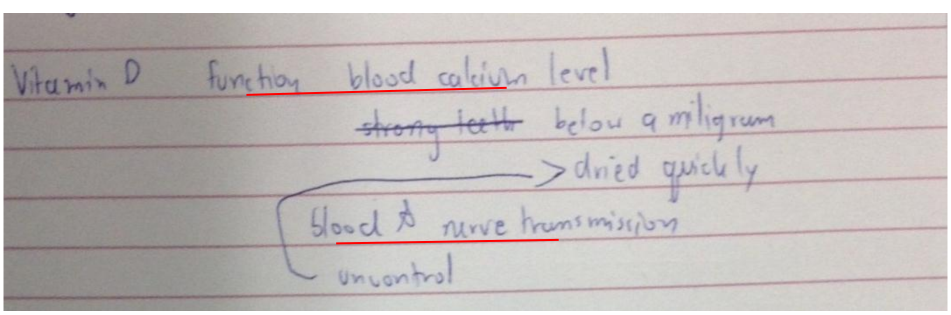
> I tried to detect key words and technical terms in health science such as nerve, muscle contraction and nerve. (P7/*Vitamin D*)

Quote 6.25

> I heard function and then main concept. I predicted that it was going to be the effect of not having enough of Vitamin D. I focused on the technical terms used to describe it. I heard 'nerve transmission'. (P2/*Vitamin D*)

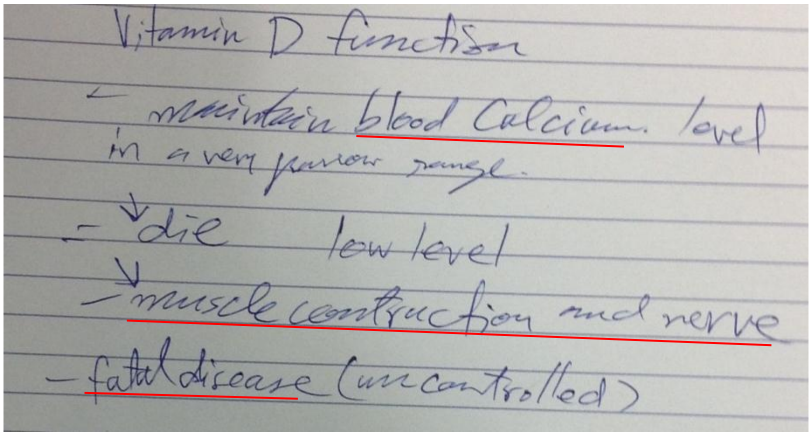
In their notes, these two participants (P7 and P2) wrote:

Note 6.8



(P2/*Vitamin D*)

Note 6.9



(P7/*Vitamin D*)

In sum, the selective attention strategy, as revealed by the stimulated recall data, was used after the participants had seen the image provided in the tasks, listened to a few sentences, and predicted the main point of the text. After predicting and contextualizing the initial input, the participants then selectively paid attention to the vocabulary they considered or predicted to be related to the main idea of the listening texts. In itself this strategy is useful since it could lead the user directly to the main point of the texts. However, in this study selective attention was used in combination with prediction, the success of which depended also on the activation of cognitive processes at the lower-level (linguistic processing). The use of this strategy thus appeared useful when text processing at the lower-level functioned effectively and the prediction was in line with the text's information, but not necessarily in other cases.

6.2.3 Directed attention

Directed attention refers to the act of paying attention to the task at hand and avoiding any distraction. It is different from selective attention in that directed attention concerns getting attention back to focusing on one or two particular aspects of a text after it has been disrupted, whereas selective attention is about maintaining focus on the input predicted to be the theme of the text. Directed attention involves noticing when one's own attention is slipping or diverting from the ongoing listening passage and redirecting it back, which is only possible in the presence of attention monitoring (Goh, 2002).

In this study, each task contained an image related to the audio text. While listening to the text, some participants thought that it was necessary to pay attention to the image in order to understand the audio text. At the same time, they also wanted to take notes of important information since they were required to produce a summary of the listening input afterwards. Listening, in this study, was therefore likely to involve multi-tasking, such as paying attention to the incoming text, processing the text for comprehension, linking the listening information to the image, and taking notes. This was potentially cognitively demanding for some participants as they had only one opportunity to listen. The analysis showed that the participants were occasionally distracted from the passage and subsequently redirected their attention. Directed attention was used with the highest number of the participants in the tasks where *Hans Krebs* was used as listening input, followed by *Corruption*, *Talent*, and *Vitamin D,* respectively.

One activity that disrupted several participants' attention while listening was note-taking. Although it was not obligatory for the participants to take notes, in practice they did a lot of note-taking. Perhaps this is because they knew that they would have to refer to some notes when constructing a summary later. While taking notes, the participants reported that they had missed taking in some information of the text. When this had happened, some stopped taking notes and paid full attention to the incoming text. This can be seen in the following quotes.

Quote 6.26

> I got confused and stopped taking notes, trying to understand how Walberg got involved in this story. I was trying to concentrate on it. (P11/*Hans Krebs*)

Quote 6.27

> I noted down his name, Hans Krebs….. at this moment I was lost. I stopped writing and was trying to focus on it. I was listening for words that I was familiar with. I was lost….. and ….I heard 'great children'. I guessed it was about inspiration. (P4/*Hans Krebs*)

Quote 6.28

> The audio was fast. I couldn't take notes while listening. I was lost when I did it. I stopped taking notes… I was concentrating on the listening. (P10/*Talent*)

Quote 6.29

> I noted down 'define', 'restricted', 'sense', and 'broad'. I was lost, here. I stopped taking notes, trying to concentrate on the listening text. (P1/*Talent*)

Quote 6.30

> I couldn't follow the listening. I missed the text when taking notes. I was lost… I heard 'workforce'. I looked at the screen. I couldn't focus. I stopped writing and listened for familiar words. (P4/*Talent*)

Directed attention was also used when the participants listened but did not understand the meaning of the text being delivered. Instead of focusing on ambiguous information, some participants directed their attention away from it and focused on the text coming next. One participant said:

Quote 6.31

> I knew it was 9 milligrams per something, but I couldn't catch it. I ignored it and focused on the incoming text. (P6/*Vitamin D*)

Another instance of the use of directed attention was when participants got lost while listening to the text; namely, when they could not catch the points/ideas in the listening message or did not know what they were listening to. They then tried to overcome this problem by directing their attention to known and familiar words. In *Corruption*, for instance, they said:

Quote 6.32

> I couldn't concentrate on the text at the beginning. I didn't understand what I was listening to. I was trying to get my attention back by listening for words I was familiar with. …. I also looked at the graph. (P1/*Corruption*)

Quote 6.33

> I didn't know what I was listening to. I heard ‘social cost’, but couldn't contextualize it. I kept on listening, trying to get familiar words. (P10/*Corruption*)

In *Hans Krebs*, the participants appeared to have the same problem; that is, they did not understand the ideas in the text. Therefore, they concentrated hard on the incoming text in order to detect known or familiar words to compensate for missing information. They, for example, indicated:

Quote 6.34

> I knew it was about a person, but I lost my understanding when it was about his father and Warburg. I knew only that it was about ‘chemical reaction’ and about science. I paid complete attention on the information delivered next to catch words I was familiar with. (P1/*Hans Krebs*)

Quote 6.35

> When I was lost, I focused on words only, trying to catch as many known words as possible. (P9/*Hans Krebs*)

In *Talent* and *Vitamin D*, the participants also directed their attention to known words/phrases when they could not grasp particular points in the text, to solve their comprehension problems. Some of them, for example, stated:

Quote 6.36

> I was not familiar with the topic. I didn't understand the text, but I tried to focus on words….. I was trying to write down the words I could catch. (P11/*Talent*)

Quote 6.37

> The story was far from my background knowledge. I heard it, but I didn't understand it. I needed more time to process the text… I was lost. I didn't know from the listening how Vitamin D is important to our body. This distracted my listening in the later part. I was trying to understand it by listening for known words. (P11/*Vitamin D*)

Although directed attention was reported being used by the highest number of the participants, it is not clear to what extent the use of this strategy facilitated their comprehension processes. One participant stated very clearly that she was not successful in her attempt when listening to *Talent*:

Quote 6.38

> I could not guess what the story would be about. I felt there was too much information to remember. I didn't know… I couldn't synthesize it. I was lost. I was trying to focus on it, but it didn't work well. (P6/*Talent*)

Of the participants who reported directing their attention back to the listening passage after losing the main point of what they were listening to, only a few explicitly reported being successful in this manner. For example, P3, who did not catch the purpose of *Hans*

Krebs at the beginning, was trying to keep up with the text until he heard ‘memoir’, which enabled him to realize that the text was describing a personal story. This participant expressed:

> *Quote 6.39*
>
> I was a bit confused with some vocabulary here, but I kept listening…I heard 'memoir'. Then I knew the text was describing a person's life. (P3/*Hans Krebs*)

To conclude, directed attention was found to be used in three different situations in order to solve comprehension problems. One is to direct attention to one activity at a time; that is, to pay full attention to the listening text and stop taking notes when the participants were not clear about what they were listening to. Another use of directed attention was to stop thinking about information that was not clear or not understood and to pay full attention to the incoming text. Lastly, and importantly, it was used when the participants lost track of what they were listening to. In that case, attention was redirected to known and familiar words, with the aim of generating ideas from recognized or known words. Successful application of this strategy was found to vary across participants. It is likely that this depended on whether they had enough linguistic knowledge to understand the upcoming text.

6.2.4 Comprehension monitoring

Comprehension monitoring is a strategy used for checking and confirming how well one understands listening materials (see 2.4.6). It involves noticing possible errors in inferences and confusion or incoherence in different parts of the interpretation. Comprehension monitoring involves checking continuous understanding of the text, checking predictions against the incoming text, and making adjustments where necessary.

To monitor their comprehension, listeners make use of both external and internal resources, including textual information, visuals, prior knowledge, and contextual information.

In the present study, comprehension monitoring was used by the highest number of participants, compared to the other metacognitive strategies. Compared across the listening topics, the highest number of participants monitoring their understanding was found for *Talent*, followed by *Corruption*, *Vitamin D* and *Hans Krebs,* respectively. Comprehension monitoring was employed for several reasons. One was to monitor ongoing understanding of the text. This was when participants checked whether they had fully understood the text or missed any points whilst listening. Below are excerpts from the participants' verbal reports.

Quote 6.40

> I understood that there were two important ideas; one is about corruption and the other was about development assistance. I knew that development assistance was used to compare with corruption. …. I was trying to connect 'Dr. Arnold' to another piece of information. OK, I got the point. (P9/*Corruption*)

Quote 6.41

> I nodded my head here because I understood what the speaker was talking about. (P1/*Talent*)

Quote 6.42

> I was thinking about the main point of the text. Yeah, it was… it was about the definitions of talent. (P3/*Talent*)

In order to assess their ongoing comprehension, some participants checked their understanding of the spoken text against the visual information. In their listening to *Corruption*, where graphic information was provided, the participants reported:

Quote 6.43

I was following the listening text, trying to create its outline. I linked it back to the graph and thought whether it described the graph…… I checked if I understood the graph correctly. (P3/*Corruption*)

Quote 6.44

I understand that corrupt payment was high. This was shown in the higher bar. The shorter bar was related to the development assistance. …OK, my understanding was tuned to the listening passage. (P7/*Corruption*)

Comprehension checks, however, did not always give a satisfying result. Some participants checked their comprehension and found that they did not understand (part of) a passage.

Quote 6.45

I understood that corruption was ten times. The figure given by the speaker explained the graph. I missed some information when I took notes. Towards the end, the listening was very fast, faster than the beginning. I lost most of it. (P10/*Corruption*)

Quote 6.46

I heard ‘ten times’ and ‘one trillion’. I turned to the graph. I didn't see ‘one trillion’ but I saw ‘corrupt payment’, ‘development assistance’, and ‘financial flows’. I was trying to link these two bars to the passage, but I couldn't. I had no background. I couldn’t connect the graph to what the speaker said. (P1/*Corruption*)

After checking their own understanding and realizing that they had missed some information or had a gap in their understanding, some participants chose to ignore it, use their background knowledge to overcome the gaps, and redirect their attention to the incoming text. As they described:

Quote 6.47

> There was some information that I didn't understand. I sometimes had to add my own opinion to connect the information I had gained. When the speaker talked about ‘regressive process’, I had no ideas what it was. I ignored it. (P6/*Corruption*)

Quote 6.48

> I was planning to catch as many words as I could. But there were several words I was not familiar with. I had to ignore them and focus on the text coming next. (P5/*Hans Krebs*)

Some participants chose to ignore what they had missed because they evaluated it as a supporting detail, not contributing to understanding the main point. They explained:

Quote 6.49

> I had lost some information here, some examples about an airline. The information was very dense and the listening was fast. I didn’t catch it all. It was OK because it was just a detail. (P6/*Talent*)

Quote 6.50

> I can't follow this point [examples of talent in airline business]. It was very fast. I thought it should be ok. I got enough words to write. I knew I had missed some details but not the main idea, I guessed. (P8/*Talent*)

Quote 6.51

> I was thinking that there might be some details that I had missed, but it didn’t matter. I got enough information to summarize. (P4/*Vitamin D*)

Quote 6.52

> I got lost when it was about a disease here [towards the end of the listening]. It was hard to follow. Too much information was given. I chose to dismiss it because I thought I got the main point. I had enough to write a summary. (P12/*Vitamin D*)

Quote 6.53

> He [the speaker] mentioned the number ‘9 milligrams’ of something, which I had missed. It was ok although we didn't get it because we were not a medical student who had to know an accurate number because if they didn’t, they would harm the patient. (P3/*Vitamin D*)

As shown in the *Quotes* 6.49-6.53, after noticing that they had missed some information, a decision was quickly made to solve the problem. These participants ignored the problematic parts because they thought those parts had little or no impact on their comprehension of the main idea. However, although some participants knew that they had missed some important information which affected their understanding, they directed their attention away from it and focused on the incoming text so as not to lose their understanding of the incoming part. They reported:

Quote 6.54

> When it was about the number and proportion, it was difficult to me. I didn’t have time to process it. I skipped it because I didn't want to miss the incoming text. I paid attention to ‘die’ instead. (P7/*Vitamin D*)

Quote 6.55

> The organization of information made me confused. There were no signal words such as first, second...It was hard to recognize. I heard 'and then' so I knew that the content moved to the second definition, and I already missed the first one. I ignored it and listened carefully to the rest. (P7/*Talent*)

In addition to assessing ongoing comprehension, the participants evaluated their own predictions. Some of them realized that they had made correct predictions, as they maintained:

Quote 6.56

> I looked at the screen. I heard ‘talent is not easy to define’. I predicted that it was about the meaning of talent. I made an accurate prediction. I predicted that first

> the passage would present definitions given by different people and then conclude by giving the speaker's definition. (P2/*Talent*)

Quote 6.57

> I knew it [the passage] was explaining the meaning of 'talent' in different ways. I divided the ideas into 1, 2 and 3. I kept listening and knew that there were two main categories of the definitions, narrow and broad. It was what I expected to hear. (P9/*Vitamin D*)

In checking predictions against the incoming information in the *Talent* and *Vitamin D* texts, the participants generally found that their prediction corresponded to the texts. This could be because these two texts had a more conventional lecture structure where the first few sentences of the texts hint at the main point and the rest of the text explains the main point stated at the beginning. *Talent* began by asking 'what is talent?' and described 'people in the business of writing tend to define talent in very different ways'. The rest of the text then described the meaning of talent in different contexts. In *Vitamin D*, similarly, the first few sentences stated very clearly that the text was about a real function of Vitamin D, and the rest of the text explained how Vitamin D works. In this way, the participants were likely to accurately predict the upcoming text even if they had only listened to the initial input.

In *Corruption* and *Hans Krebs*, however, the main point of the texts was not explicitly stated at the beginning of the passage and the participants rarely accurately anticipated what they were going to hear. They therefore had to verify their prediction in order not to misunderstand the main point of these listening passages. Some participants, although they did notice that their prediction was wrong, did not manage to verify it in time to keep up with the key point of the text. P8, for instance, reported:

Quote 6.58

> I knew what I predicted was wrong according to the listening text, but it was late. The main point had already gone. (P8/*Corruption*)

Most inaccurate predictions occurred in the *Hans Krebs* task. The participants, for example, expressed:

Quote 6.59

> The passage was not relevant to my prediction. I anticipated that I would hear more about his life, what he did, and what he produced, but I didn't. Until it was the end of the text, I realized that I had missed the main point. (P9/*Hans Krebs*)

Quote 6.60

> I felt that the flow of the story was disrupted. I was confused about the focus of the text. I took notes of his work, but I thought the focus was shifted from his work to his family. It was not about his work or his biography…Here, the speaker said, 'all kinds of obstacles'. I was lost. (P10/*Han Krebs*)

Quote 6.61

> In the beginning, it was that this man was a role model. But then the theme was changed. I was lost and could not follow it. It was like we were driving a car very fast and we didn't see the turn when we had to. By the time that we realized it, we had gone too far to go back. I knew I missed the key point. (P11/*Hans Krebs*)

In particular, the analysis shows that the participants needed to realize that the prediction was not in line with the listening passage in order to verify their prediction and understand the main idea correctly. Most participants started off their listening with a prediction which was based on the initial input and their background/topical knowledge. Consequently, they listened with some anticipated information in mind. When the listening input was not presented in a traditional text structure, where the main point is mentioned at the beginning of the text and the details are given further in the text, the prediction appeared to be wrong. The participants misunderstood the main point unless they were able to notice discourse markers used to present the important information of the texts. They had to realize that the text did not follow their prediction, and adjust it in time in order to understand the main point of the text. This was the case in *Hans Krebs*, where a large number of the participants relied on their prediction and did not realize that their prediction was wrong. Although some participants (P3, P9, P10, and P11) noticed the text's content was not structured in the way that they had predicted, only one participant (P3) managed to verify his prediction and divert his attention just in time to focus on the main point of the text. This participant said:

Quote 6.62

> In the beginning, it was about his name and his work [Krebs' cycle]. I thought it was going to be about his biography. Then the speaker said 'wonderful example .....' When it came near the end, the speaker concluded that 'no matter what we say to them...'. I thought this confirmed that my prediction was wrong. The text focused on 'he is a wonderful example of scientist who overcame all kinds of obstacles'. What was said in the beginning was just an introduction, but not the main point. To me, this text was not for a science class but a psychology class. (P3/*Hans Krebs*)

In addition to assessing the main point and checking and verifying predictions, comprehension monitoring was used to monitor participants' note-taking. When realizing that their note-taking behaviour was not effective, two participants adjusted their note-taking strategies. They reported:

Quote 6.63

> I was trying to summarize and create a coherent text while listening, but it didn't work. I lost some ideas, so I used bullet points instead. (P8/*Corruption*)

Quote 6.64

> I was very slow in taking notes. Whenever I wrote down, I missed the beginning part of the next chuck of information. I knew I had missed some important points

> about ‘chemical reaction’. I then tried to mentally remember the information instead of writing it down. (P2/*Hans Krebs*)

To summarize, comprehension monitoring is one of the useful metacognitive strategies used by the participants in order to help them listen successfully. It was used mainly to continuously assess ongoing comprehension and evaluate whether the missing information would affect overall understanding of the text. An action that was executed alongside the comprehension check was ignoring the missing information and redirecting attention to the incoming text. Another role of comprehension monitoring was to check the incoming text against predictions and verify it when the prediction deviated from the listening text. However, to achieve this goal the participants had to pay close attention to the incoming text. This was difficult since in some texts where the discourse pattern was not familiar to the participants (in particular *Hans Krebs*), the participants thought their particular prediction was accurate (when, in fact, it was not) and remained attached to it. They thus misunderstood the main point of the text, unless they had verified their prediction in time to keep their focus on the main point. Some participants also monitored their note-taking behaviour when finding out that it did not work well and was causing problems in comprehension.

6.2.5 Real-time assessment of input

Real-time assessment of input was originally introduced in the literature on L2 listening strategies by Goh (2002). It involves making on-the-spot decisions about whether a particular part of the input is necessary for task completion. It is different from decision-making in comprehension monitoring in that in the process of real-time assessment of input, the listeners make a decision to either foreground or background a set of information. The listeners may choose to pay more attention to a set of information when they regard it as important for task achievement or ignore some pieces of information when they disregard them as crucial. During comprehension monitoring, however, decision-making involves deciding whether unfamiliar words or ideas will affect one's understanding of the whole text. If not, listeners may choose to ignore these and redirect their attention to the task at hand, not being disrupted by a problematic understanding.

In this study, the analysis showed that real-time assessing of input was used mainly to consider whether a set of information would explain the main point of the story. When the participants considered that it was not likely to, they tended to ignore it and pay attention to other parts of the text that they thought would contribute to the main point. Four participants were found using real-time assessment of input when listening to *Vitamin D*. Two used it for *Corruption*, and one each for *Talent* and *Hans Krebs*.

In *Vitamin D*, where the highest number of the participants reported applying real-time assessing of input, the participants assessed and disregarded 'Vitamin D maintains strong bones and teeth' as a necessary piece of information. This is because the text that followed it described that 'it does that by accident'. The participants thus directed their attention to 'the main function of Vitamin D' which they thought was the main point of this text. They indicated:

Quote 6.65

> I didn't focus on this [maintaining strong teeth]. He [the speaker] said 'it does that by accident'. I don't think it was a main point……I heard 'blood clotting'. I didn't focus on it. There was too much information. (P8/*Vitamin D*)

Quote 6.66

> I ignored this part 'maintain strong bones and teeth' because he said ‘it does by accident’. I tried to catch what the main function was if that was just an accident. (P3/*Vitamin D*)

Some participants ignored pieces of information because they felt that the passage contained too many details for them to remember and thought they had got enough information to produce a summary. They said:

Quote 6.67

> I focused only on the main point. It was about the benefits of Vitamin D. I didn't focus on the disease. I thought it was a small detail. (P5/*Vitamin D*)

Quote 6.68

> I think there were a lot details and a lot of technical terms. Each term came with very long modifiers such as blood calcium level, 9 milligrams per 100 milliliters. I did not pay attention to them. I didn’t think they were the focus of the text. (P6/*Vitamin D*)

In *Corruption*, despite the participants’ lack of understanding of some pieces of information, they chose to foreground these instead of decreasing their importance. This was because they felt that they were important pieces of information:

Quote 6.69

> I couldn’t link '10 times' to the graph……I was a bit confused here. I thought it was an important piece of information. I included it (10 times) in my summary anyway. I regarded it as one of the key ideas. (P4/*Corruption*)

Quote 6.70

> I think the idea about ‘the poor’ was interesting. The speaker first mentioned it in the beginning of the passage. Then he mentioned it again. I considered it as one of the key ideas and included it in my summary. (P5/*Corruption*)

When listening to *Hans Krebs*, one participant ignored the information about Krebs' work because he felt that the text focused more on how Krebs became successful despite a lack of parental support than on his famous piece of work. The participant indicated:

> *Quote 6.71*
>
> I knew that the passage emphasized the idea that people can be successful though they lack parental support. It was not about how he became famous. ….The main point started when he said 'despite his father discouraged him ……' I wrote his name because I needed it for my summary, but I didn't pay attention to 'at the peak of his career'. I thought it was only an example. (P3/*Hans Krebs*)

In *Talent*, one participant appeared to ignore a set of information, i.e. examples given by the speaker, and waited for the main point she had anticipated to hear. She indicated:

> *Quote 6.72*
>
> When he (the speaker) talked about how people defined talent in different businesses, I listened, but didn't take notes. It wasn't important for the summary. I was waiting for the speaker's definition of talent, which I thought was the main idea. (P2/*Talent*)

In general, the application of real-time assessment of input was useful to the participants. The listening passages contained several ideas, some of which were details and examples given by the speaker to clarify key points in the text. The decision to foreground or background information seemed to help decrease cognitive load and maintain focus on the main point of the passage. Real-time assessment of input therefore seems to be one of the strategies contributing to success in text comprehension.

6.2.6 Comprehension evaluation

Comprehension evaluation refers to the final check of a listener's own textual interpretation for accuracy, completeness, or acceptability. Comprehension evaluation is not the same as comprehension monitoring, although both involve assessing the correctness of what has been understood (Goh, 2002). Comprehension evaluation takes place towards the end or after an individual has finished listening to the complete input and has arrived at some interpretation. Comprehension monitoring, on the other hand, is the ongoing process of checking whether comprehension is taking place during listening and maintaining focus on the incoming text.

In the present study, three participants reported using comprehension evaluation. All three used it in their listening to *Corruption* and one also used it in *Vitamin D*. However, no participants reported evaluating their overall comprehension in *Talent* and *Hans Krebs*. Comprehension evaluation, as revealed by the stimulated recall data, was mainly activated when the participants felt that they had not caught or recognized enough ideas while listening to the text. Thus, towards the end of the listening or after they had finished listening, they self-evaluated their understanding in order to rehearse what they had comprehended. They stated:

Quote 6.73

> I understood the text just a little bit. I knew that it was about corruption. I didn't know what was emphasized in the text. (P5/*Corruption*)

Quote 6.74

> I saw the graph. I didn't understand the text at the beginning, but I gained some ideas at the end of the listening text, when it was about taxes and income. (P11/*Corruption*)

Quote 6.75

> I understood that Vitamin D was important to our body, but I didn't understand it when the speaker talked about the process and when it was about maintaining blood calcium, blood calcium level, and its effect. (P4/*Vitamin D*)

After self-evaluating his own understanding of the text and finding that he did not understand most of the ideas, one participant (P4) reported applying two cognitive strategies to complete the task – elaboration and reconstruction. That is, he brought in his topical knowledge to elaborate his understanding of the text and reconstructed a summary based on his elaborated information and words/phrases he had noted down. He indicated:

> *Quote 6.76*
>
> I listened. I took notes, but I did not really catch the points of the text. I had to create my story based on these words [the words in the notes] and my background knowledge. (P4/ *Vitamin D*)

To summarize, comprehension evaluation was not used by many of the participants. It was basically used when they thought they had missed key ideas of the text or did not catch as much information as they had expected to, to complete the task. After self-evaluating and seeing gaps in their own understanding, cognitive strategies such as elaboration and reconstruction were activated to solve comprehension problems and to complete the tasks. As it helped the participants to notice their comprehension gaps and activate other strategies that can help compensate for the gaps, this strategy is thus considered a useful strategy.

6.3 Summary

Six metacognitive strategies were reported being used by the participants to oversee and manage their listening process, i.e., 1) pre-listening comprehension, 2) selective attention, 3) directed attention, 4) comprehension monitoring, 5) real-time assessment of input, and 6) comprehension evaluation. Pre-listening preparation was applied in the beginning of the listening task, before the audio files started. It was mainly used to rehearse the task instructions, set listening goals, and lower anxiety. Comprehension evaluation, where the participants assessed their understanding of the entire text, was used towards the end of the listening task. These two strategies appeared to be used by a small number of the participants, compared to the other metacognitive strategies found.

Four metacognitive strategies – selective attention, direct attention, comprehension monitoring, and real-time assessment of input – were evident in the dataset and used to manage ongoing listening processes. Selective attention was used in combination with the cognitive strategy of prediction. That is, after the participants had listened to the initial part of the text, they started to predict its theme and anticipate the upcoming text. As a consequence, their attention was then selectively directed to what they expected to hear. Selective attention appeared to be counterproductive when predictions were not in line with the content of the listening passage, unless the participants managed to verify their prediction and align it with the text.

Comprehension monitoring, which was used to check ongoing understanding of the passage, to assess predictions and verify their accuracy, and to evaluate note-taking strategies, was employed by a large number of the participants. They monitored their comprehension on the basis of the information available during the listening, including visual information, the audio text, and background knowledge. What was crucial in comprehension monitoring, as found in this study, was to check predictions against the incoming text and to verify it in time when the prediction was not in agreement with the listening text. To achieve this, the participants had to pay close attention to the linguistic information and be able to notice a shift in text genre or discourse structure, if there was any. Otherwise, they could easily misunderstand the main point of the text, as was the case in *Hans Krebs*, where many participants misunderstood the main point of the text because they were not aware that their prediction was wrong.

The findings also showed that for comprehension monitoring to work effectively, two other metacognitive strategies, i.e. real-time assessment of input and directed attention, had to be activated properly. After checking for comprehension, the participants appeared to notice some gaps in their understanding. They then considered whether it was a main point or a detail. If they thought the information they had missed was the main point of the listening text, the participants attempted to figure it out immediately. This was done by paying close attention to the incoming text, in order not to miss the point. If it was not the key idea, they ignored it and redirected their attention to the incoming text.

Chapter 7 Listening comprehension processing behaviours compared between tasks with different modalities and performance levels

7.1 Introduction

In this chapter, sub-analyses of the data reported on in the previous chapters (4-6) are presented. More specifically, this chapter compares the participants' cognitive and strategic processing behaviours between different performance levels and between the listening-to-speak and listening-to-write tasks (RQ 1a). The chapter begins by presenting the content scores of the participants whose cognitive and strategic processing behaviours were compared (7.2). Next, the results of the comparisons are presented in three sections, according to the processing behavior studied, i.e. cognitive processes in 7.3.1, cognitive strategies in 7.3.2, and metacognitive strategies in 7.3.3. In each of these sections comparisons are made both between tasks with different modalities and between performance levels. Section 7.4 compares overall comprehension processing behaviours between intermediate- and highly-able listeners during different periods of the listening tasks. The final section (7.5) summarizes the results of all comparisons.

7.2 Participants' performance scores

Table 7.1 presents the content scores of the 12 participants (P1-P12) whose task processing behaviours were investigated. Although, as discussed in section 3.7.3, the task performances were scored on different criteria (content, pronunciation, and fluency in the listening-to-speak tasks; and content, vocabulary, and grammar in the listening-to-write tasks), only the content scores for the oral/written summaries have been compared for the present purpose. This scoring criterion was chosen because it is typically taken to be an indicator of the test-taker's input comprehension ability in integrated test tasks. Adopting the PTE academic scoring descriptors, the participants' content summaries were marked on a three-band scale (0, 1, or 2). Summaries which provided the main point accurately and also all relevant supporting details of the listening passage were considered to be good and complete summaries, and 2 (a full mark) was allotted on the content. Summaries providing a fair summary of the text (missing one or two aspects of content) were allocated 1 out of 2. Summaries which failed to include main points or relevant details delivered in the listening passage were assigned 0. The marking was done by two experienced PTE academic raters, and their scores were averaged in case of band differences. Only two raters were used in this study because of limited financial resources since the two raters were paid to score, as is usual.

Based on their content scores, the participants were categorized into three groups: high-, average-, and low-scoring participants. Participants with a content score of 2 were considered high-scoring participants for that topic. Participants whose content score ranged between 1.0-1.50 were regarded as average-scoring participants, and those with a score of 0-0.5 on content were referred to as low-scoring participants for that topic.

Participant	Listening topics/Scores								Total (8)
	Corruption (0-2*)		Talent (0-2*)		Vitamin D (0-2*)		Hans Krebs (0-2*)		
	L-S	L-W	L-S	L-W	L-S	L-W	L-S	L-W	
P1	.50		1.00			1.50		.50	3.50 (43.75%)
P2	.50		1.00			1.00		.50	3.00 (37.5%)
P3	2.00		2.00			2.00		2.00	8.00 (100%)
P4		1.00		1.00	1.00		.50		3.50 (43.75%)
P5		2.00		1.00	1.50		1.00		5.50 (68.75%)
P6		.50		1.00	1.50		.00		3.00 (37.5%)
P7	1.00			.50	1.00			.00	2.50 (31.25%)
P8	1.50			2.00	2.00			.00	5.50 (68.75%)
P9	1.00			.50	1.00			1.50	4.00 (62.5%)
P10		1.00	1.00			2.00	.00		4.00 (50.0%)
P11		1.50	1.00			1.50	1.00		5.00 (62.5%)
P12		1.50	.00			1.50	.00		3.00 (37.5%)

**The scores on summary content were averaged between the two raters.*
L-S: Listening-to-speak tasks, L-W: Listening-to-write tasks

Table 7.1: Participants' performance scores

As Table 7.1 shows, differences in content scores can be observed between the four listening tasks and the performance of each individual participant varied across the listening topics. P8, for instance, was evaluated to be a high-scoring participant in the tasks with *Talent* and *Vitamin D* as the listening input, an average-scoring participant in *Corruption*, and a low-scoring participant in *Hans Krebs*. P10 was a high scorer in *Vitamin D*, an average one in *Corruption* and *Talent*, and a low scorer in *Hans Krebs*. As the content scores of almost all participants appeared to vary between the tasks, it is difficult to profile individual participants as high-, average-, or low-scoring in general (with the exception of P3, who was awarded the top content score on each task). Therefore, in order to investigate whether there were any differences in the listening processing engaged in by participants of different performance levels and in different tasks, different participants were contrasted according to their content score on individual tasks. Although the aim was to compare a high-scoring participant's processing with that of a low-scoring participant, for some tasks there was no high-or low-scoring participant, in those cases, an average-scoring participant's data was used.

Table 7.2 shows which participants' processing was contrasted for each task. In the listening-to-speak tasks, the processing of P3 was compared to that of P2 in the task with *Corruption*, P3 to P12 in *Talent*, P8 to P7 in *Vitamin D*, and P5 to P12 in *Hans Krebs*. With regard to the listening-to-write tasks, P5 was compared to P6 in *Corruption*, P8 to P7 in *Talent*, P3 to P2 in *Vitamin D*, and P3 to P7 in *Hans Krebs*. As there were no low-scoring participants in both the listening-to-speak and listening-to-write tasks with *Vitamin D*, a comparison was made between the high- and average-scoring participants. In the *Hans Krebs* listening-to-speak task, a comparison was made between the average- and the low-scoring participants. Given that only two performances were compared each time, the comparisons were made on limited data. However, as the comparisons were made four times in listening-to-speak and in listening-to-write tasks, it is hoped they provide patterns of processing behaviours by participants with different performance levels in different tasks.

Listening topics	**Listening-to-speak**			**Listening-to-write**		
	High-scorer	**Average-scorer**	**Low-scorer**	**High-scorer**	**Average-scorer**	**Low-scorer**
Corruption	P3		P2	P5		P6
Talent	P3		P12	P8		P7
Vitamin D	P8	P7		P3	P2	
Hans Krebs		P5	P12	P3		P7

Table 7.2: The participants whose listening behaviours were compared for each listening topic and between tasks with different modalities

7.3 Comprehension processing behaviours compared between performance levels and tasks with different modalities

For the purpose of the comparisons, sub-analyses were conducted on the stimulated recall data, the participants' notes, and their content summaries. Specifically, the cognitive processes and strategies adopted by the 12 participants with different scores and for tasks with different modalities were compared. The results are presented as follows, according to the three aspects investigated, namely cognitive processes, cognitive strategies, and metacognitive strategies.

7.3.1 Cognitive processes

As presented in Chapter 4, six cognitive processing types, i.e. acoustic-phonetic processing, word recognition, parsing, semantic processing at local and discourse levels, and pragmatic processing, were adopted by different numbers of the participants in comprehending the listening input in the listening-to-summarize tasks. This section compares the processes activated by the participants with different performance levels and for different tasks.

When comparing between performance levels (see Table 7.3), it was found that the high-scoring participants performed both the lower-and the higher-level processes. The participants whose content summary of *Corruption* and *Talent* was assigned a full mark (2) (in both the listening-to-speak and the listening-to-write tasks) activated both semantic processing at the global level and pragmatic processing. The low scorers, on the other hand, engaged only in the lower-level processes, i.e. acoustic-phonetic processing, word recognition, and parsing.

Listening-to-speak Tasks	Acoustic-phoneic processing	Word recognization	Parsing	Semantic processing (local)	Semantic processing (global)	Pragmatic processing
High-scorer: Corruption (2*)	✓	✓	✓	✓	✓	✓
High-scorer: Talent (2)	✓	✓	✓	✓	✓	✓
High-scorer: Vitamin D (2)	✓	✓	✓	✓	✓	
Average-scorer: Hans Krebs (1-1.5)	✓	✓	✓	✓	✓	
Average-scorer: Corruption (1-1.5)	✓	✓	✓	✓		
Average-scorer: Talent (1-1.5)	✓	✓	✓	✓		
Average-scorer: Vitamin D (1-1.5)	✓	✓	✓	✓		
Low-scorer: Hans Krebs (0-0.5)	✓	✓				
Low-scorer: Talent (0-0.5)	✓	✓	✓			
Low-scorer: Corruption (0-0.5)	✓	✓	✓			

Listening-to-write Tasks	Acoustic-phoneic processing	Word recognization	Parsing	Semantic processing (local)	Semantic processing (global)	Pragmatic processing
High-scorer: Corruption (2)	✓	✓	✓	✓	✓	✓
High-scorer: Talent (2)	✓	✓	✓	✓	✓	✓
High-scorer: Vitamin D (2)	✓	✓	✓	✓	✓	
High-scorer: Hans Krebs (2)	✓	✓	✓	✓	✓	✓
Average-scorer: Hans Krebs (1-1.5)	✓	✓	✓	✓		
Average-scorer: Corruption (1-1.5)	✓	✓	✓	✓		
Average-scorer: Talent (1-1.5)	✓	✓	✓	✓		
Average-scorer: Vitamin D (1-1.5)	✓	✓	✓	✓		
Low-scorer: Hans Krebs (0-0.5)	✓	✓	✓			
Low-scorer: Corruption (0-0.5)	✓	✓	✓			

*content score

Table 7.3: The cognitive processes compared between performance levels and between tasks with different modalities

In *Vitamin D*, the results show that the participants with a full mark (2) on content engaged in semantic processing at the discourse level but not in pragmatic processing, and no participants scored low (0-0.5) on the content. This indicates that the participants in general were able to process the text at linguistic processing and comprehension processing levels effectively, and that it was not necessary to rely on contextual information to determine the main point of the text. In *Corruption*, *Talent* and *Hans Krebs*, the participants receiving a full mark (2) on the content relied on both semantic processing at the global level and pragmatic processing to understand the texts. This might suggest that in these texts the main point was not as explicitly indicated as it was in *Vitamin D*, and the participants therefore had to rely on contextual information in order to completely understand the texts. For the *Hans Krebs* passage, only one participant received a full mark (2) on the content. This might suggest that, compared to the other listening topics, *Hans Krebs* was linguistically the most difficult and only one participant managed to process the text at the higher-level successfully to understand the text's main point.

The comparison of the cognitive processes performed in the tasks with different response modalities – speaking and writing – in general did not point to any distinctive differences in the cognitive processes adopted. As shown in Table 7.3, the participants with a full content score (2) in both the listening-to-speak and the listening-to-write tasks engaged in high-level processes, i.e. semantic and pragmatic processing. The low-scoring participants in the two modalities, on the contrary, engaged only in low-level processes, i.e. acoustic-phonetic processing, word recognition, and parsing. There was no evidence that these participants engaged in the higher-level processes.

To sum up, the results showed that the cognitive processes adopted by the participants varied between the listening topics. The high-scorers who achieved a full content score in each task engaged in both lower-and higher-level processes. The low scorers, however, were found to have engaged only in lower-level processes. Within the same performance levels, different cognitive processes were found to be activated when compared across listening passages. In *Vitamin D*, the participants achieved the highest score (2) although they did not report engaging in pragmatic processing, whereas in *Hans Krebs*, they did not achieve it when they did not report engaging in pragmatic processing. This might be because in *Vitamin D*, the main point of the text was clearly stated at the beginning of the listening passage, whereas in *Hans Krebs* the main idea was not explicitly indicated but had to be inferred by the participants. Interestingly, the participants who reported engaging in only lower-level cognitive processing scored low on the content (0-0.5). One factor that seemed to stimulate diversity in cognitive processing behaviours was the nature of the input materials.

7.3.2 Cognitive strategies

This section compares the cognitive strategy use between performance levels and between tasks with different modalities. For this purpose, the frequencies of each cognitive strategy used by the participants (whose processing behaviours were compared for each listening topic, as shown in Table 7.2) were counted.

The results show that there are considerable differences in the cognitive strategies used when compared between the performance levels but not between tasks with different modalities. Figures 7.1-7.4 illustrate the proportion of the cognitive strategies used in the four listening passages, i.e., *Corruption*, *Talent*, *Vitamin D*, and *Hans Krebs*, respectively. In each figure, four pie charts demonstrate the cognitive strategies employed. The two charts at the top present the strategies performed in the listening-to-speak task by the highest and the lowest-scoring participants, while the bottom two present the strategies used by those in the listening-to-write task.

In the tasks where *Corruption* was used as listening input, the results show that the low scorers in both tasks (the listening-to-speak and the listening-to-write tasks) reported using reconstruction strategies (11%-13%) whereas the high-scorers did not report doing so (see Figure 7.1). The participants all reported using inferencing, elaboration, and fixation. However, the extent of use of inferencing by the high-scorers was about twice as much as that reported by the low scorers in both tasks. On the other hand, in the listening-to-write modus, the low-scoring participant applied fixation about twice as much as the high-scoring participant.

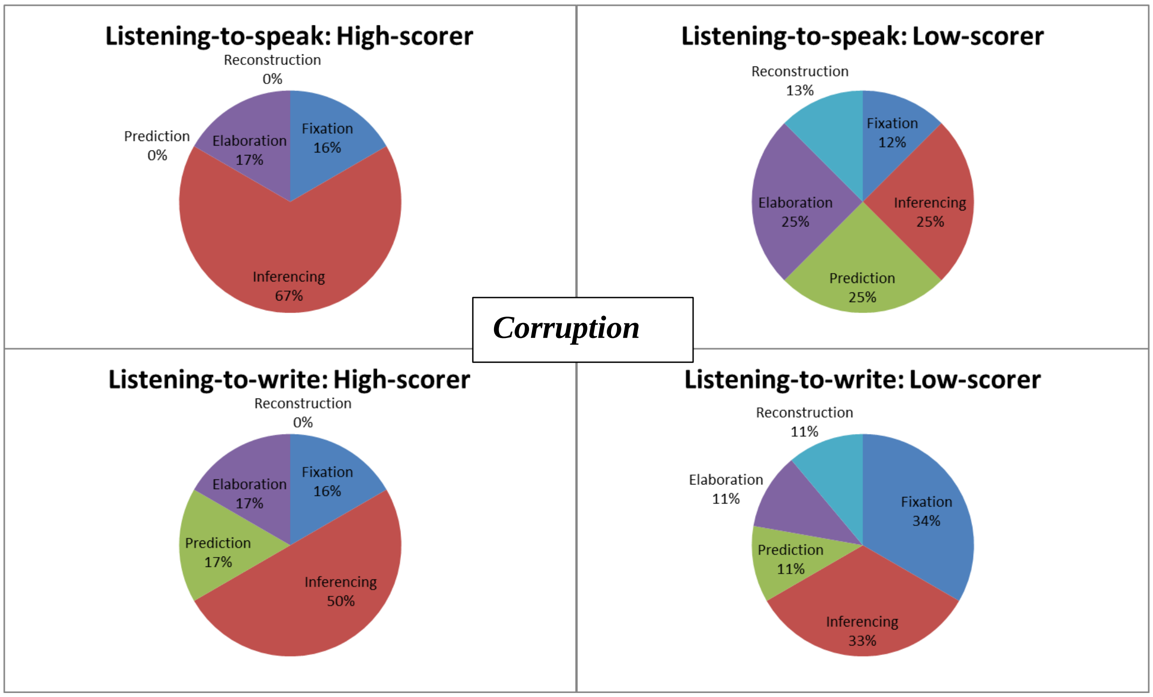


Figure 7.1: Cognitive strategies compared between tasks with different modalities and between performance levels (*Corruption*)

With respect to *Corruption*, it can be concluded that the low scorers in both listening-to-speak and listening-to-write tasks employed more types of cognitive strategies than the high-scorers. Reconstruction for instance was used by the low scorers but not by the high-scorers. However, the high-scorers reported using inferencing more frequently than the low scorers in both tasks.

In the tasks where *Talent* was used (see Figure 7.2), the results showed that only two cognitive strategies – prediction and inferencing – were used by the high-scoring participants, and their use of inferencing was about four times higher than the use of prediction in both tasks (86% vs.14% in the listening-to-speak and 83% vs. 17% in the listening-to-write tasks). The low-scoring participants reported using more types of cognitive strategies than the high scorers. In addition to inferencing and prediction, they used fixation and elaboration. In the listening-to-speak task, fixation was the most frequently used strategy, comprising 43% of the entire cognitive strategy use. In the listening-to-write task, fixation and inferencing were the two cognitive strategies the most frequently used by the low scorer, each accounting for 43%. In all cases, there was no report of reconstruction use in *Talent*.

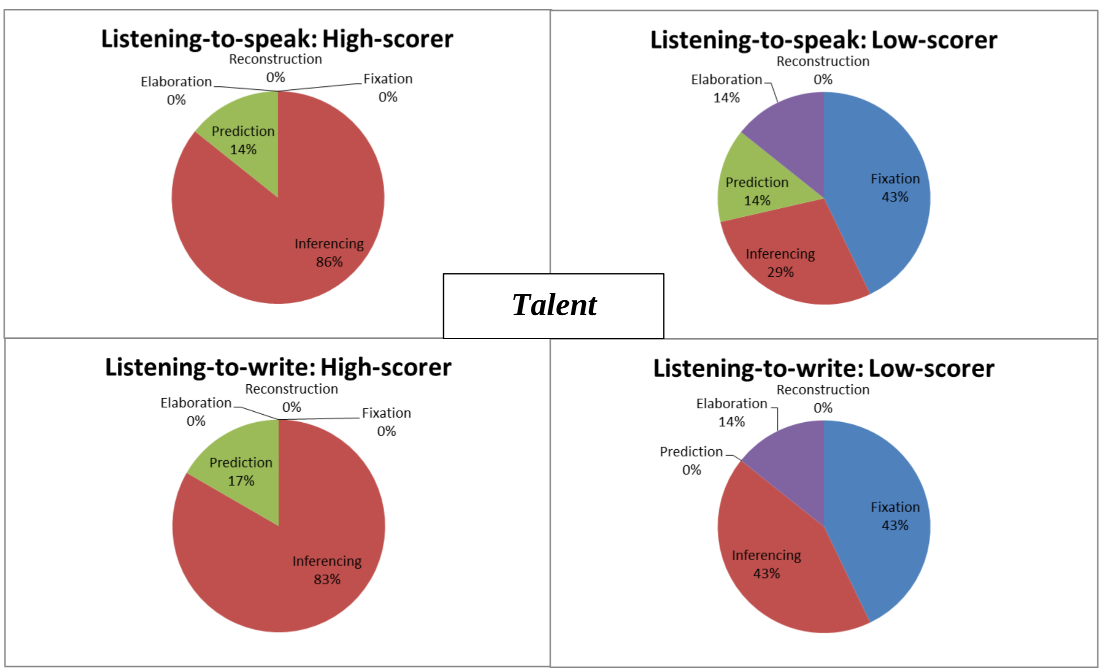


Figure 7.2: Cognitive strategies compared between tasks with different modalities and between performance levels (*Talent*)

In conclusion, the comparison of cognitive strategy use in the tasks with *Talent* in both the listening-to-speak task and listening-to-write task shows quite a similar pattern.

More types of cognitive strategies were reported being used by the low scorers in their listening to *Talent* in both tasks. While the high-scoring participants used inferencing most often, the low-scoring participants used fixation and inferencing the most frequently. The proportion of inferencing used by the low scorer was considerably lower than that used by the high scorers.

In *Vitamin D*, there were no low-scoring scorers, so the comparison was made between high- and average- scoring participants. The results showed that prediction and inferencing were the two cognitive strategies used by the participants in both tasks and between the performance levels (see Figure 7.3). The use of inferencing was, however, about two to three times higher than the use of prediction in both task types. Elaboration appeared to be used only by the average scorers. Fixation was only reported in the listening-to-write task. Both elaboration and fixation were however used far less than inferencing in both tasks.

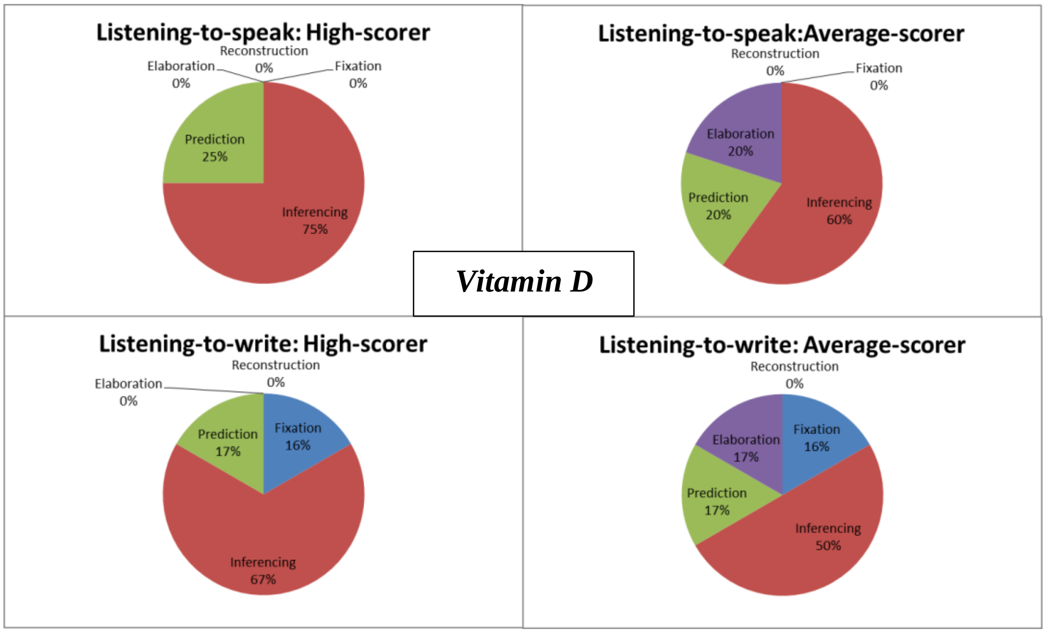


Figure 7.3: Cognitive strategies compared between tasks with different modalities and between performance levels (*Vitamin D*)

In short, the results showed that the participants relied mainly on inferences to fill gaps in their comprehension while listening to *Vitamin D*. Although other cognitive strategies (prediction, elaboration, and fixation) were used, they accounted for a smaller proportion of the strategies. The only difference found when comparing the strategy used between tasks with different modalities is that fixation was used in the listening-to-write task but not in listening-to-speak tasks. However, it only accounts for a small number.

In the tasks with *Hans Krebs*, the results showed that the participants activated a number of cognitive strategies. High- and average- scoring participants in the listening-to-speak and the listening-to-write tasks were quite similar in that they employed inferencing the most frequently (50%) and their use of prediction accounted for a small proportion (17%) in both tasks. In the listening-to-speak task, in particular, the single average-scorer reported using fixation in 16% of the cases and used reconstruction in

17% of the total reporting of cognitive strategy use. This participant however did not report using any elaboration. In the listening-to-write task, the high scorer did not appear to use any fixation or reconstruction strategies, but used elaboration in 33% of the cognitive strategy use cases. With regard to the low scorers, they reported using inferencing least (14-17%). In the listening-to-speak task, elaboration was a highly used strategy by the low scorer, accounting for 43%. In the listening-to-write task, the two most used strategies were reconstruction and fixation (33% each).

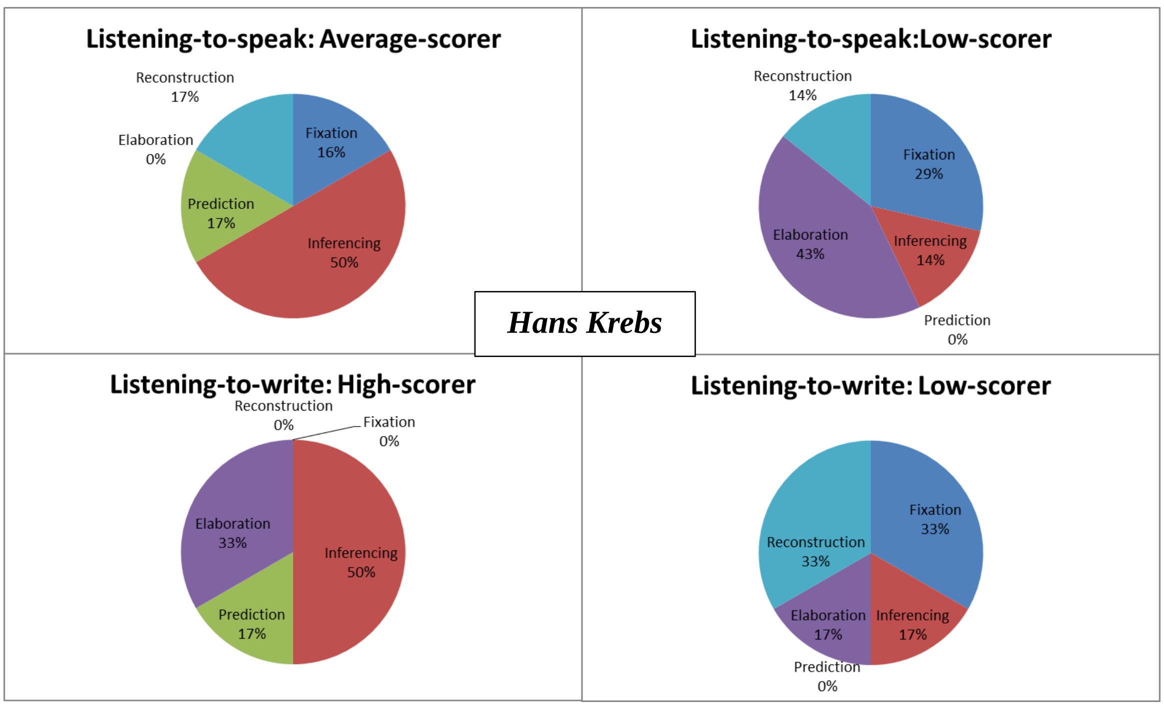


Figure 7.4: Cognitive strategies compared between tasks with different modalities and between performance levels (*Hans Krebs*)

It can be summarized that in *Hans Krebs*, quite similar strategies were reported by the high-, average-, and low-scoring participants to understand the text. However, the strategies were used with different ranges of frequency. While the high- and average-scoring participants tended to rely more on inferencing, the low-scoring participants depended more on elaboration, fixation, and reconstruction. The strategies employed to complete tasks with different modalities were quite similar in types and frequency.

In conclusion, the comparison of the cognitive strategy use shows that the participants differed considerably when comparing between the performance levels. To comprehend the listening input, the high scorers depended considerably on inferencing since they reported using this strategy the most frequently for all the listening passages. In *Corruption*, *Talent* and *Hans Krebs,* in particular, the proportion of inferencing used by the high scorers was about twice as high as that used by the low scorers. On the other hand, the low scorers were found to have relied mainly on three cognitive strategies: fixation, reconstruction, and elaboration, to solve their comprehension problems. When comparing between the tasks with different modalities (listening-to-speak versus listening-to-write), it was found that quite similar strategies were activated although in some listening passages they were used to different extents.

7.3.3 Metacognitive strategies

As presented in Chapter 6, six metacognitive strategies were used to manage the listening process in the listening-to-summarize tasks. They are 1) preparing for listening, 2) selective attention, 3) directed attention, 4) comprehension monitoring, 5) real-time assessment of input, and 6) comprehension evaluation. This section compares the use of these strategies between performance levels and tasks with different modalities. The results of the comparison are presented in Figures 7.5-7.8, beginning with the tasks with *Corruption* and then those with *Talent*, *Vitamin D,* and *Hans Krebs*, respectively. In each figure, the two pie charts at the top present the metacognitive strategies used in the listening-to-speak tasks, comparing these between the participants with different performance scores. The two charts at the bottom compare those used in the listening-to-write tasks.

With reference to *Corruption* (see Figure 7.5), the results show that comprehension monitoring and preparing for listening are the two metacognitive strategies used by all the participants compared. Comprehension monitoring was, however, used the most frequently by the high scorer in the listening-to-speak task, accounting for 45% of the participant's use of metacognitive strategies. Selective attention was another metacognitive strategy frequently used by this participant, accounting for 33%. The low scorers in both the listening-to-speak and the listening-to-write tasks appeared to use directed attention the most frequently. The proportion of this strategy used in the listening-to-write task was almost twice as much as that used in the listening-to-speak task (60% vs. 33%). Only the high-scorer in the listening-to-write task appeared to use comprehension evaluation and real-time assessment of input and only the high scorer in the listening-to-speak task did not report using directed attention.

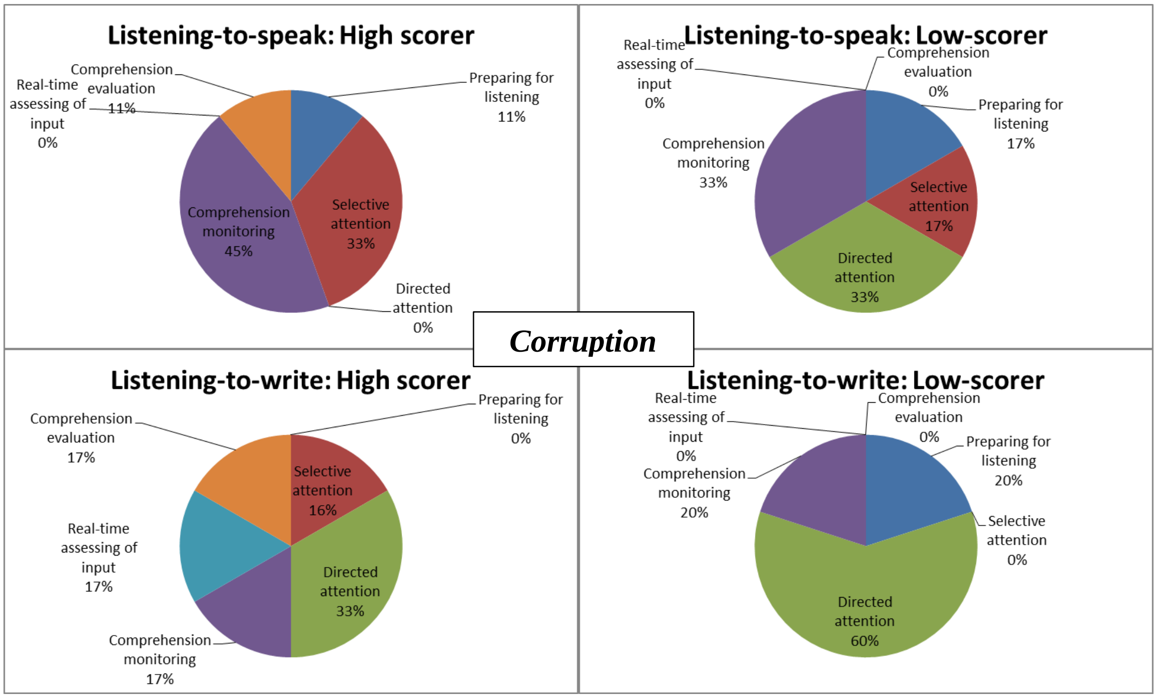


Figure 7.5: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Corruption*)

In short, the results suggest that both the high and the low scorers are relatively different in their use of metacognitive strategies while listening to *Corruption.* The high scorers appeared to have employed more types of metacognitive strategies than the low scorers in both tasks. The two strategies frequently used in both the tasks and by the high and the low scorers are comprehension monitoring and directed attention. They, however, were used at different frequency rates. Comparing between the tasks with different modalities, quite similar types of strategies were found and used with similar extents.

Figure 7.6 illustrates the metacognitive strategies activated in the tasks with *Talent*. Only three metacognitive strategies were reported for these tasks, i.e. comprehension monitoring, selective attention, and directed attention. The high and the low scorers, except for the low scorer in the listening-to-speak task, reported using comprehension monitoring the most frequently, comprising approximately two-thirds of all metacognitive strategy use. In the listening-to-speak task, comprehension monitoring was used in combination with selective attention whereas in the listening-to-write task it was used with directed attention. With regard to the low-scoring participant in the listening-to-speak task, although he reported using comprehension monitoring and directed attention, his use of directed attention was about two times higher than that of comprehension monitoring (67% vs. 33%).

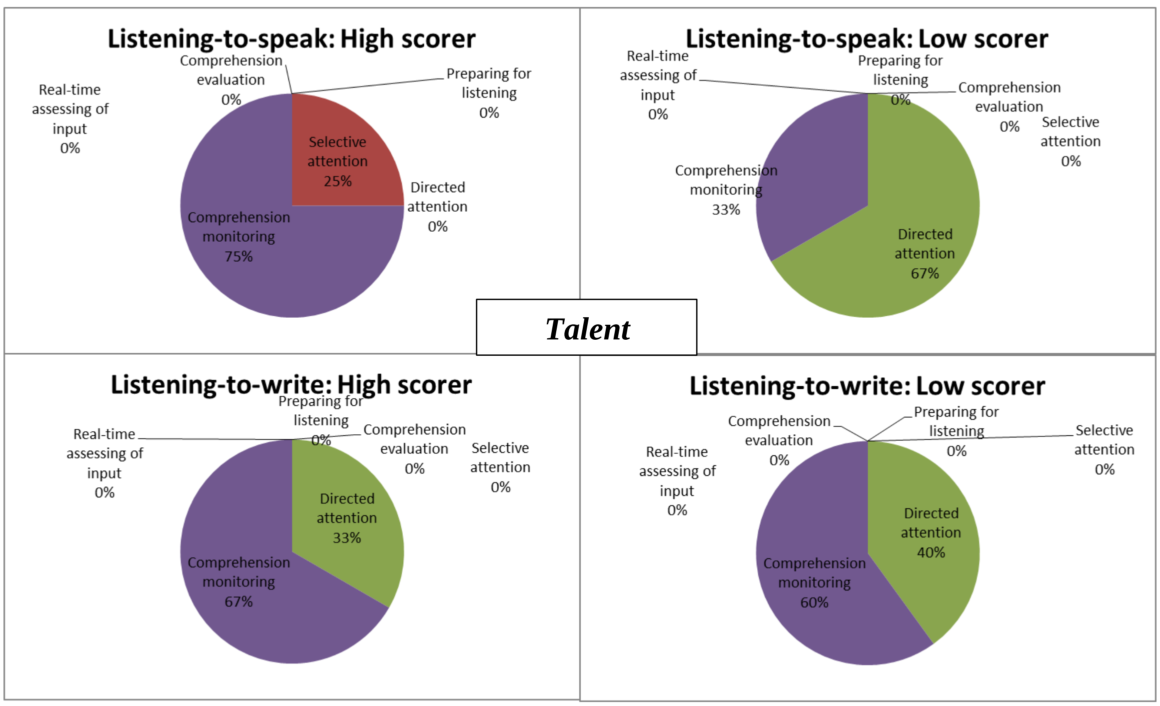


Figure 7.6: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Talent*)

To sum up, three metacognitive strategies (comprehension monitoring, selective attention, and directed attention) were reported being used by the participants to manage their listening to *Talent*. Comprehension monitoring was the most frequently used by the high-scoring participants in both tasks and the low scorer in the listening-to-write task. The low scorer in the listening-to-speak task relied mainly on directed attention.

In the tasks with *Vitamin D*, only high and average scorers were found and as a result these were compared. A variety of metacognitive strategies was reported being used by these participants, with the high-scorer in the listening-to-speak using more types of metacognitive strategies than the others. The two common metacognitive strategies shared among the participants are comprehension monitoring and selective attention. Comprehension monitoring was used the most frequently by the participants in both the listening-to-speak and the listening-to-write tasks. Interestingly, only high-scoring participants in both tasks reported using real-time assessment of the input (17% and 20%).

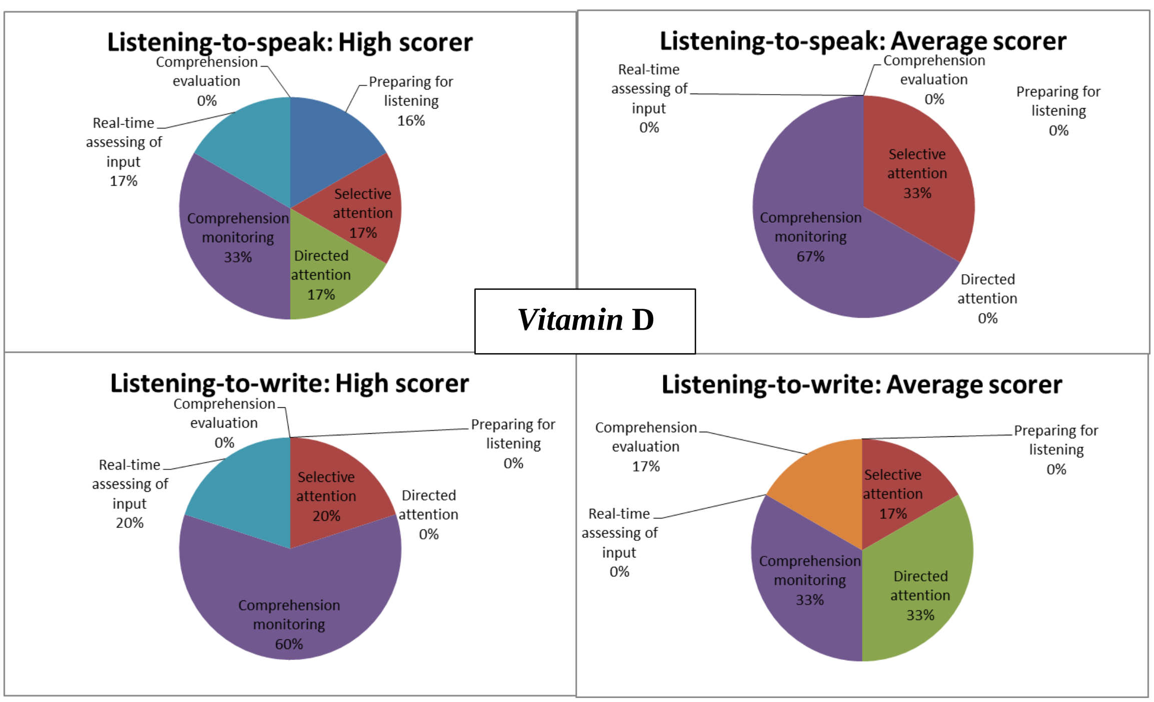


Figure 7.7: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Vitamin D*)

In summary, the results suggest that the participants, including high and average scorers, are similar in their metacognitive strategy behavior in that they relied mainly on comprehension monitoring to manage their listening to *Vitamin D*. However, they differed in that the average scorers did not appear to have engaged in real-time assessment of input, or assessed whether the input was important for the summary task, whereas the high scorers did. A slight difference was found when comparing between the tasks with different modalities. That is, comprehension evaluation was used in the listening-to-write task but not in the listening-to-speak task. This however accounts for a small number.

In the tasks with *Hans Krebs*, the results showed that the high/average versus low scorers differed in their use of metacognitive strategies (see Figure 7.8). The high and the average scorers appeared to use more types of strategies than the low scorers in both tasks. The single average scorer in the listening-to-speak task reported using comprehension monitoring with the largest proportion (60%). The high scorer in the listening-to-write task, on the other hand, reported using real-time assessment of input the most frequently (45%) and his use of comprehension monitoring accounted for only 11%. For the low scorers, directed attention comprised the largest proportion, accounting for 67% in the listening-to-speak task and 100% in the listening-to-write task; in fact, it was the only metacognitive strategy reported by the low scorer.

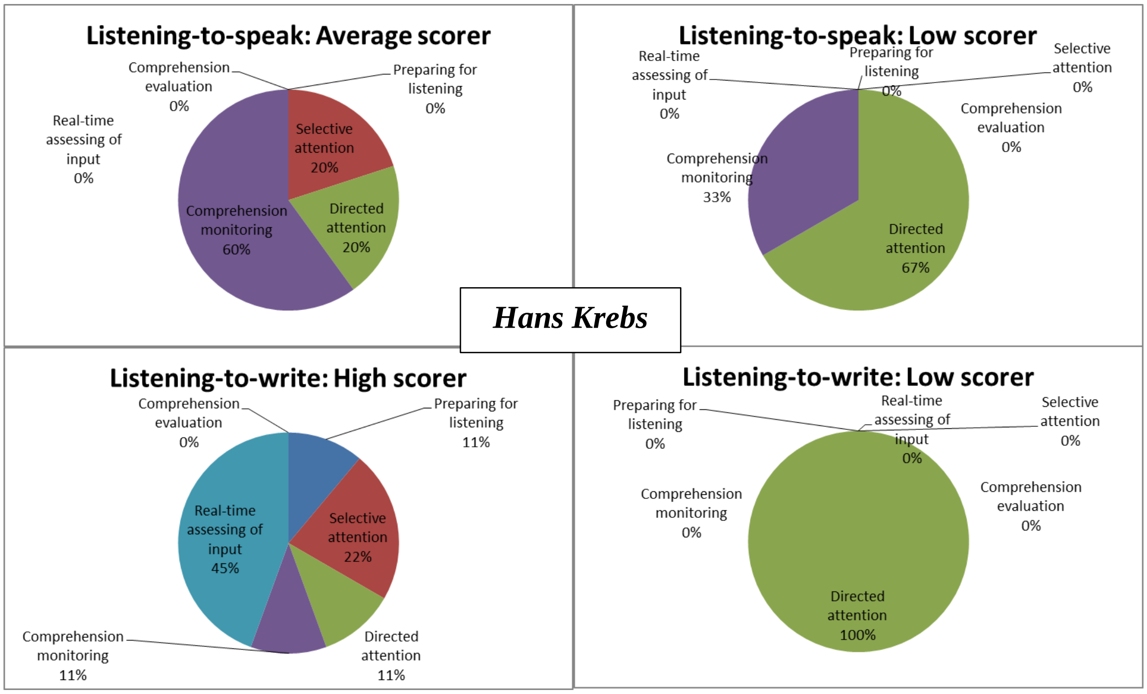


Figure 7.8: Metacognitive strategies compared between tasks with different modalities and between performance levels (*Hans Krebs*)

In sum, real-time assessment of input is the metacognitive strategy most frequently used by the high scorer, and comprehension monitoring by the average scorer in their listening to *Hans Krebs*. The low scorers on the other hand relied on directed attention. This might be because during the listening process, their attention was frequently disrupted and they thus had to direct their attention back to the listening text. Quite similar types of strategies were activated in both listening-to-speak and listening-to-write tasks.

To conclude, the comparison of metacognitive strategies used by participants with different performance levels revealed a number of differences. The high and the average scorers were likely to use more types of metacognitive strategies with a high frequency of comprehension monitoring use. For some listening passages (*Hans Krebs* and

Corruption), the high scorers also reported using real-time assessment of input. Their use of this strategy was in fact the main difference between the metacognitive strategies used by the high and the lower scoring participants. With regard to the low scorers, apart from applying fewer types of metacognitive strategies, they relied heavily on the directed attention strategy. This can be clearly seen in their listening to *Hans Krebs*, *Talent*, and *Corruption*. The comparison between the tasks with different modalities (listening-to-speak and listening-to-write) indicated that quite similar metacognitive strategies were used although they were used to different extents.

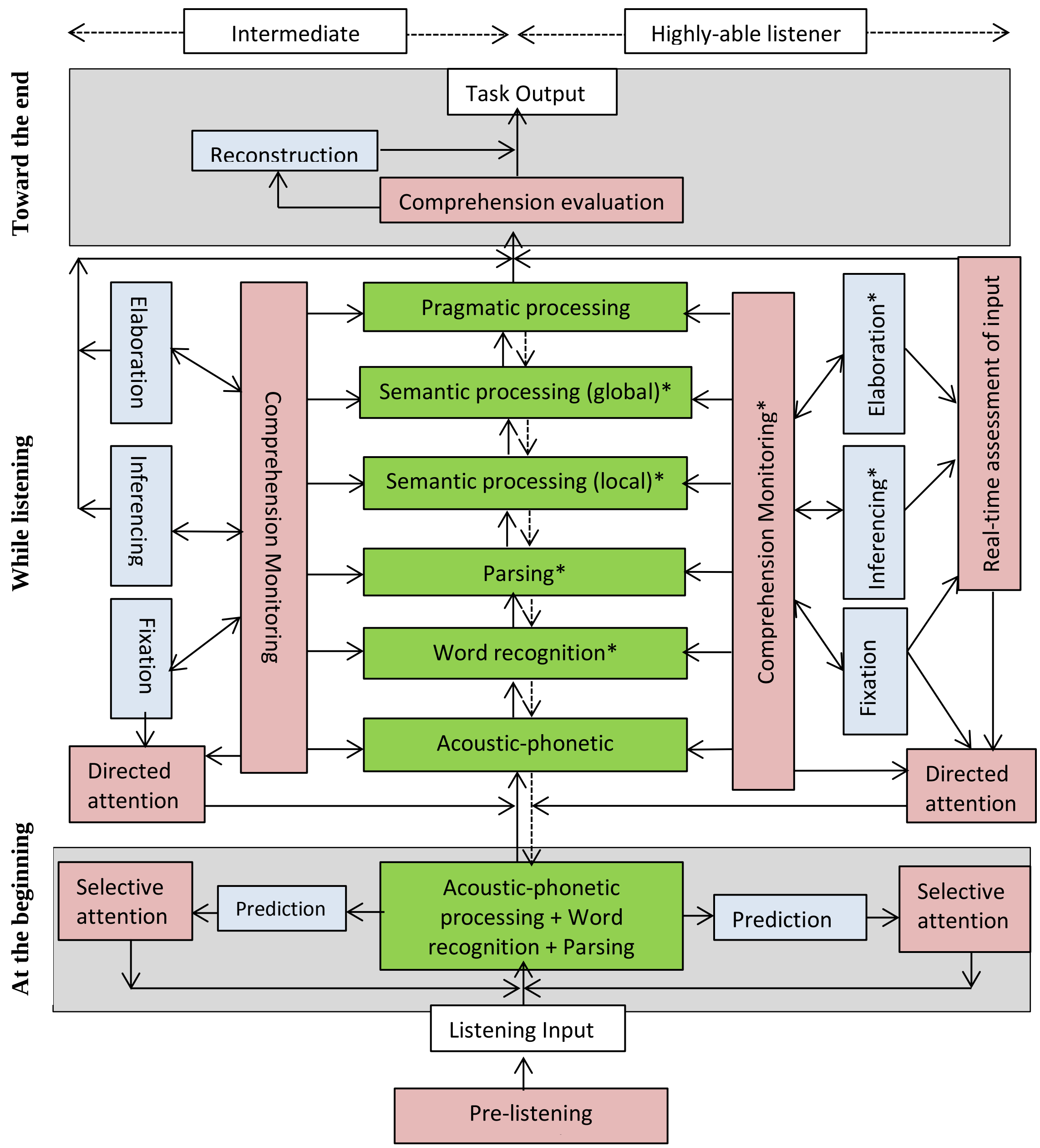
7.4 Overall picture of comprehension processing behaviours compared between intermediate and highly-able listeners during different periods of the listening tasks

The comparisons of the participants' listening comprehension processing behaviours suggests that the processes and strategies activated by the participants were different when compared between the performance levels, and a few small differences were observed when compared between the tasks with different modalities (listening-to-speak vs. listening-to-write). To obtain a clearer picture of how the participants with different performance levels differ in their task processing, the study further analyzed their comprehension processing behaviours at different periods of the listening: 1) at the beginning, 2) while listening, and 3) towards the end of the listening.

For this purpose, the data from the two participants who scored the highest (P3) and the lowest (P7) overall were further analyzed and compared (see Table 7.1 for the participants' scores). The participant scoring the highest was considered a highly-able listener in this study as he obtained full marks for content on all tasks (8/8), whereas the one who scored the lowest (2.5/8) can be regarded as an intermediate listener rather than a low ability listener, since this participant had 8.0 in the IELTS listening band and an overall band of 7.0. As the comparisons of the cognitive processes and strategies of different participants (see Sections 7.2 and 7.3) indicate quite similar patterns in the strategies used by participants at the same performance level (high/average vs. low), only the processes and strategies of the two 'extreme' participants, i.e., the highest and the lowest scorers, are compared here.

This analysis showed that, in general, the types of cognitive processes and strategies activated by these two participants were quite similar. Nevertheless, four main differences were observed: 1) the intermediate listener did not appear to have performed the cognitive processes as effectively as did the highly-able listener, 2) the intermediate listener did not appear to use real-time assessment of input whereas the highly-able listener did, 3) the intermediate listener's use of the cognitive strategies (inferencing and elaboration) was not as successful as that of the highly-able listener, and 4) the intermediate listener reported using the reconstruction strategy whereas the highly-able listener did not.

Figure 7.9 visualizes the overall comprehension processing behaviours of the highly-able listener compared to the intermediate one. The listening process has been divided into three periods, i.e. the beginning of the listening, while listening, and towards the end of the listening. At the centre of the figure are the cognitive processes activated by both intermediate and highly-able listeners.



***These processes/strategies are activated more efficiently by the highly-able listener.**

Figure 7.9: Comprehension processing behaviours compared between the highly-able and intermediate listeners during different periods in the listening tasks

On the left-hand side of the figure are the strategies (cognitive and metacognitive) employed by the intermediate listener to assist the listening process. Those activated by the highly-able listener are presented on the right-hand side.

At the beginning of the listening

The analysis showed that the processes activated at the beginning of the listening process by both the intermediate and highly-able listeners were similar. That is, they both prepared for the listening by rehearsing the task instructions, reducing their task anxiety, and getting ready to listen. After listening to a few sentences and being able to recognize some words or chunks of words, they started anticipating what they were going to hear next. Then they focused their attention selectively on what they predicted to hear.

While listening

A major difference between the intermediate and the highly-able listeners' comprehension processing was found whilst they were listening and in their abilities to process text at the lower level and the use of some strategies – inferencing, elaboration, and real-time assessment of input. After predicting the incoming text and selectively focusing their attention on what they had anticipated to hear, the listeners continued processing the text. The cognitive processes were activated interactively. Input from each stage of processing appeared to be either passed onto the next stage of processing or sent back for further processing when it was ambiguous or unknown. At the same time, the incoming text was processed and informed by the results of earlier and on-going processing. What monitored these cognitive processes was comprehension monitoring. However, the cognitive processing of the intermediate listener was not as efficient as that of the highly-able listener, especially the word recognition and parsing processes, where several key words/ ideas went unrecognized by the intermediate listener.

Along with cognitive processing, three cognitive strategies (fixation, inferencing, and elaboration) were employed to solve listening problems. Fixation in particular was used at the linguistic processing level to identify words and phrases that were unclear, both in meaning and spelling. Then attention was directed away from the problematic part of the information and brought back to the input signal through the use of metacognitive strategies (directed attention). At the comprehension processing level, two cognitive strategies (inferencing and elaboration) were employed to conceptualize the meaning of the text. The inferences and the elaborations made by the intermediate listener were, however, not as effective as those of the highly-able listener. This was because, as revealed by the data obtained, the intermediate listener did not make inferences on key information, and the elaborations were based on background knowledge that was not in line with the textual information, resulting in the participant's misunderstanding of the text. Some of the fixed, inferred, and elaborated information was then sent back to assist cognitive processing whereas some was passed on for further processing.

In the case of the highly-able listener, after making inferences and elaborations, he also checked the inferred or elaborated information against the incoming text and assessed whether it was relevant to the current text and whether that piece of the text was important for the summary task. When the highly-able listener thought that the text was important but it was not entirely clear how it would connect to his current textual understanding, he appeared to rely on bottom-up processing to solve his problem. That is, he directed his attention to the incoming text and paid attention to words/phrases or parsed information in order to figure out the links between the texts' ideas. Unlike the highly-able listener, the intermediate one adhered to his inferred and elaborated information and did not appear to question or assess whether that information was accurate according to the text.

Towards the end of the listening

The analysis revealed that towards the end of the listening, both the highly-able and the intermediate listener evaluated their overall comprehension of the text. The highly-able listener appeared to be satisfied with his own understanding and thought that he was ready to produce a summary of the listening. The intermediate listener, on the other hand, realized that he had missed some key points of the listening text. Therefore, he reconstructed the missing ideas from the information he had gained in order to make his understanding of the text complete.

7.5 Summary

In this chapter, sub-analyses were conducted on the data on the cognitive processes and strategies employed by the participants to understand the listening materials in the listening-to-summarize tasks. This was done to investigate whether there were any differences in the processes and strategies between performance levels and tasks with different modalities. The results suggest only a few small differences in the listening comprehension processing behaviours when compared between the tasks with different modalities (the listening-to-speak versus listening-to-write tasks). That is, the participants reported activating quite similar types of processes and strategies with slightly different levels of frequency. However, comparing between the performance levels, more differences were observed. With regard to the cognitive processes, it was found that the low scorers were likely to perform linguistic processing and semantic processing at the local level. Although they conducted semantic processing at the global level, it was not as successful as that performed by the high-scorers. With reference to their use of the cognitive strategies, the low-scoring participants appeared to employ more types of strategies than did the high-scoring participants. While the high-scorers appeared to rely more on inferencing, the low scorers appeared to more frequently use elaboration, fixation, and reconstruction. In terms of metacognitive strategies, the high-scoring participants appeared to activate more types of these strategies. In addition, the high scorers appeared to rely more on comprehension monitoring and real-time assessment of input, whereas the low scorers employed directed attention more frequently. When participants used similar processing/strategies, often there were differences in the extent of usage and degree of success between the high and low scorers.

Chapter 8 Perceptions of tasks and of listening task difficulty

8.1 Introduction

Test-takers' perceptions of tasks and task difficulty are another element that this research aimed to study, in addition to the comprehension processes and strategies used to complete the tasks. This is to see how test-takers perceive the tasks in terms of their authenticity, fairness, and difficulty. The second research question thus asks 'what are ESL test-takers' perceptions of (adapted) PTE Academic listening-to-summarize tasks and of listening task difficulty?' In order to answer this question, the perception questionnaires with a five-point Likert scale were administered to the 60 students participating in Group B (see 3.3) after they finished each listening-to-summarize task.

In this chapter, the participants' average performance scores on different components are first presented to give the readers insight into the level the participants achieved on the tasks (8.2). In order to evaluate the quality of the questionnaire data, the questionnaire was analysed; the results of the analysis are provided in 8.3. Next, the participants' perceptions on three aspects are described in three different sections, i.e. perceptions of the task's authenticity in 8.4, test fairness in 8.5, and listening task difficulty in 8.6. Then, the relationships between the perceptions of listening task difficulty and task performance are presented in 8.7. A summary of this chapter is provided in 8.8.

8.2 The participants’ performance scores

This section provides the performance scores of the participants whose questionnaire data were analysed and correlated to their task performances. The scores (see Table 8.1) are presented according to the tasks with different modalities and the four listening passages provided in each task group.

Task/ Listening passage	**N**	**Performance scores**											
		Content (0-2)				**Pronunciation (0-5)**				**Fluency (0-5)**			
Listening-to-speak		**Min.**	**Max.**	**Mean**	**Std.**	**Min.**	**Max.**	**Mean**	**Std.**	**Min.**	**Max.**	**Mean**	**Std.**
Corruption	30	0.50	1.50	0.92	0.37	1.00	4.50	2.35	0.67	1.00	4.00	2.60	0.64
Hans Krebs	30	0.00	2.00	0.75	0.45	1.50	3.00	2.33	0.42	1.50	4.00	2.52	0.62
Talent	30	0.00	2.00	1.03	0.51	2.00	3.50	2.45	0.51	2.00	4.00	2.85	0.53
Vitamin D	30	0.00	2.00	1.03	0.51	2.00	3.50	2.45	0.51	2.00	4.00	2.85	0.53
Listening-to-write		**Content (0-2)**				**Grammar (0-2)**				**Vocabulary (0-2)**			
		Min.	**Max.**	**Mean**	**Std.**	**Min.**	**Max.**	**Mean**	**Std.**	**Min.**	**Max.**	**Mean**	**Std.**
Corruption	30	0.50	2.00	0.97	0.41	0.00	2.00	0.92	0.47	0.50	2.00	1.00	0.47
Hans Krebs	30	0.00	2.00	0.43	0.68	0.00	2.00	0.90	0.40	0.50	2.00	1.05	0.48
Talent	30	0.00	2.00	1.12	0.49	0.50	1.50	0.98	0.28	0.00	2.00	0.93	0.41
Vitamin D	30	0.00	2.00	1.22	0.47	0.50	2.00	0.98	0.33	0.50	2.00	1.17	0.40

Table 8.1: Group B participants’ performance scores

Generally, the analyses show higher mean scores for the tasks with *Talent* and *Vitamin D* than those with *Corruption* and *Hans Krebs* in almost every component, including content, pronunciation, and fluency in the listening-to-speak tasks, and content and grammar in the listening-to-write tasks. The only exception is for the vocabulary aspect in the listening-to-write tasks where a higher mean score was found for the tasks with *Corruption* and *Hans Krebs* than for *Talent*. The overall performance scores fall at the moderate level in almost every component. The only exception was the mean content score of the listening-to-write task with *Hans Krebs* (M= 0.43), which can be considered a low performance. In the listening-to-speak tasks, the average content scores for the four listening passages ranged between 0.75-1.03 out of 2, the pronunciation average scores ranged between 2.33-2.45 out of 5, and the fluency average scores were 2.52-2.85 out of

5. In the listening-to-write tasks, the mean content scores ranged between 0.43-1.22 out of 2, the mean grammar scores were 0.90-0.98 out of 2, and the mean vocabulary scores were 0.93-1.17 out of 2.

8.3 The analysis of the questionnaires

The questionnaire, which was composed of 23 statements, aimed to measure perceptions in three respects, i.e. perceptions of 1) task authenticity, 2) task fairness, and 3) listening task difficulty (see 3.4.4). As multiple items were included to measure each component, a factor analysis was carried out to evaluate the association of the questionnaire items underlying the same construct. Table 8.2 presents the results of the analysis, according to the two types of questionnaires used: the listening-to-speak and the listening-to-write questionnaires. These two questionnaires are generally similar in terms of what they aimed to measure. The only difference was that some items were reworded to correspond to the productive skills involved in the task performance (i.e. from 'speak' to 'write').

Variable	Questionnaire	No. of items	Item no.	Eigenvalue	% of variance explained	α*
Tasks' authenticity	Listen-to-speak Listen-to-write	2	1,2	1.19 1.29	59.68 64.62	.74 .76
Tasks' fairness	Listen-to-speak Listen-to-write	3	3,4,5	2.07 1.82	69.01 60.92	.86 .81
Code complexity	Listen-to-speak Listen-to-write	5	6,7,8,9,10	1.88 1.63	37.65 32.72	.69 .67
Cognitive familiarity	Listen-to-speak Listen-to-write	4	11,12,13,14	1.08 1.68	27.02 32.94	.70 .74
Cognitive processing	Listen-to-speak Listen-to-write	5	15,16,17,18, 19	1.30 1.20	26.06 24.06	.65 .67
Communication stress	Listen-to-speak Listen-to-write	4	20,21,22, 23	1.48 1.56	37.06 39.02	.68 .69

**Reliability (Cronbach's Alpha)*
N =120

Table 8.2: The analyses of questionnaire components and reliability

To evaluate the questionnaires, two types of values were calculated, i.e. eigenvalues and reliability values. The eigenvalues are used to explain the extent to which sub-components indicate the substantive importance of a component. According to Field (2005), an eigenvalue which is greater than 1 shows that the sub-components (eigenvectors) are significantly related and cluster as a factor (variable). The eigenvalues of 1.08-2.07 found in this study can thus be interpreted to indicate that the items included in each component significantly represent the component they belong to.

In addition, the questionnaires were also analysed for their reliability. The reliability values, as indicated by α (Cronbach's Alpha), were found to vary between 0.65 and 0.86. Although generally the reliability value of $\geq$.80 is suggested to indicate good reliability, acceptable values, as suggested by Field (2005) can be lower, depending on the number of the items underlying the same construct. When a small number of items are used (e.g., 3-5 items), the acceptable value can be levelled down to 0.6 (Field, 2005). The values between 0.65 and 0.86 of the variables, which include 2-5 items in this study, thus suggested acceptable reliability of the questionnaires.

8.4 Perceptions of task authenticity

Listening-to-summarize tasks have been integrated in language testing to form more authentic test tasks and tap into the abilities required in real-world communication. In practice, however, it is largely unknown to what extent test-takers, who are directly affected by this task type, view the tasks as authentic. The first two items of the questionnaires were thus designed to tap into perceptions of task authenticity.

The results, as presented in Table 8.3, shows that a majority of the participants thought that the listening-to-summarize tasks, both the listening-to-speak and the listening-to-write, represent the tasks they encounter in their actual academic context. 63.3% of the participants agreed and 21.7% strongly agreed that the listening-to-speak tasks simulate academic situations. Only 2.5% disagreed with this. For the listening-to-write tasks, 60% of agreement and 30% of strong agreement was found, and only 5.8% disagreed with this statement. In addition, three quarters of the participants (75%) (strongly) agreed that the listening-to-speak tasks assess the English ability required in academic studies and a majority of the participants (80.8%) (strongly) agreed with this statement with respect to the listening-to-write tasks.

Tasks authenticity	**Task type**	**N**	**Frequency (percentage)**				
			Strongly disagree	**Disagree**	**neutral**	**agree**	**Strongly disagree**
1. The task simulates a situation in academic contexts.	L-S	120	0	3 (2.5%)	15 (12.5%	76 (63.3%)	26 (21.7%)
	L-W		1 (0.8%)	6 (5.0%)	5 (4.2%)	72 (60.0%)	36 (30.0%)
2. The task assessed the English ability required for academic study	L-S	120	0	5 (4.2%)	25 (20.8%)	67 (55.8%)	23 (19.2%)
	L-W		1 (0.8%)	4 (3.3%)	18 (15.0%)	67 (55.8%)	30 (25.0%)

Table 8.3: Perceptions of task authenticity

In summary, the results suggest that the participants perceive both the listening-to-speak and the listening-to-write tasks as authentic. They thought that the tasks simulated the situations they encounter in their real life and that the tasks assess the ability in English that they need to perform in academic studies.

8.5 Perceptions of test fairness

Table 8.4, which includes three statements, reports on participants' view of task fairness. The results showed that more than half of the participants (strongly) agreed that both tasks fairly assess their academic English (see Item 3), i.e. 60.8% for the listening-to-speak tasks and 66.7% for the listening-to-write tasks. When being asked whether the tasks accurately tap into their academic English listening ability, a significant number of the participants (79.1% for listening-to-speak and 73.4% for listening-to-write tasks) (strongly) agreed that the tasks were able to do so and only a very small number of the participants disagreed with this statement (5.8% for the listening-to-speak and 1.7% for the listening-to-write tasks). Regarding the potential of the tasks to accurately assess productive skills, i.e. speaking and writing, about two-thirds of the participants (strongly) agreed that the tasks accurately measured the productive skills involved in task performance, 70.9% for assessing speaking ability and 65.8% for writing ability. A small number of the participants (about 10%) thought that the tasks (both listening-to-speak and listening-to-write tasks) did not accurately assess their productive skills and about a quarter of the participants had a neutral view on this statement.

Tasks fairness	**Task type**	**N**	**Frequency (percentage)**				
			Strongly disagree	**Disagree**	**neutral**	**agree**	**Strongly disagree**
3. This is a fair way to assess my ability to use English in academic contexts	L-S	120	2 (1.7%)	10 (8.3%)	35 (29.2%)	54 (45.0%)	19 (15.8%)
	L-W		0	11 (9.2%)	29 (24.2%)	56 (46.7%)	24 (20.0%)
4. The task accurately reflects my English listening ability	L-S	120	1 (0.8%)	6 (5.0%)	18 (15.0%)	73 (60.8%)	22 (18.3%)
	L-W		0	2 (1.7%)	30 (25.0%)	62 (51.7%)	26 (21.7%)
5. The task accurately reflects my English speaking/writing ability.	L-S	119*	0	12 (10%)	22 (18.3%)	65 (54.2%)	20 (16.7%)
	L-W	120	2 (1.7%)	9 (7.5%)	30 (25.0%)	64 (53.3%)	15 (12.5%)

**1 missing*

Table 8.4: Perceptions of tasks fairness

To sum up, the participants generally thought that the tasks accurately assessed their academic English abilities and the skills involved in task performance. The number of participants who thought that the tasks accurately measured their comprehension skills (listening) was slightly higher than those considering that the tasks accurately assessed their productive skills (speaking/writing).

8.6 Perceptions of listening task difficulty

The questions on perceptions of difficulty in this study were designed to tap into the difficulty of the listening passages used as input materials. The results from this section were then related to participants' listening task performances in order to investigate their relationship. Following Skehan (1996; 1998), the perceptions of difficulty were investigated according to three input characteristics: code complexity, cognitive complexity in two areas (cognitive familiarity and cognitive processing demands), and communication stress (see 2.5.2).

8.6.1 Code complexity

Perceptions of code or linguistic complexity were investigated with reference to five features of texts' linguistic characteristics. They are lexical complexity (Item 6), syntactic complexity (Item 7), information density (Item 8), information redundancy (Item 9), and discourse complexity (Item 10). Tables 8.5-8.8 present the participants' perceptions of code complexity of the four listening passages used, i.e. *Corruption*, *Hans Krebs*, *Talent*, and *Vitamin D*, respectively in the listening-to-speak and the listening-to-write tasks.

Compared across the five aspects of code complexity, information density (Item 8) was one element (strongly) agreed on by a high number of the participants. Many (57.6%-76.7%) thought they had to process many important ideas whilst listening to three of the listening passages, namely *Corruption*, *Hans Krebs*, and *Talent*. For *Vitamin D*, a different picture was found as a third of the participants did not agree with this statement.

The perceptions of lexical difficulty, structural complexity, information redundancy, and discourse complexity were found to be similar in pattern for the tasks with *Talent* and *Vitamin D*. That is, the highest number of the participants (almost 50% for *Talent* and slightly over 50% for *Vitamin D*) disagreed that the vocabulary was difficult (Item 6) or the sentence structures were complicated (Item 7). Moreover, more than half of them agreed that there was textual redundancy and the key ideas in the texts were paraphrased or repeated more than once (Item 9). Approximately two-thirds agreed that the ideas in the passage were clearly connected (Item 10). For *Corruption*, although about half of the participants (strongly) agreed on the occurrence of textual redundancy and clear organization of the text, almost half of them held neutral views on its lexical and structural complexity.

Unlike its counterparts, the *Hans Krebs* passage was perceived to be lexically and structurally complex by almost half of the participants. Over two-thirds of them, in addition, perceived the textual information as dense and only one-fourth thought that its important ideas were paraphrased or repeated more than once. Almost half of the participants held a neutral view on this passage's discourse complexity, neither agreeing nor disagreeing that the text's ideas were clearly connected.

Code complexity (*Corruption*)	Task type	N	Frequency (percentage)				
			Strongly disagree	disagree	neutral	agree	Strongly disagree
6. Vocabulary in the listening passage was difficult for me.	L-S	30	1 (3.3%)	6 (20.0%	13 43.3%)	8 (26.7%)	2 (6.7%)
	L-W	30	1 (3.3%)	9 (30.0%)	11 (36.7%)	8 (26.7%)	1 (3.3%)
7. Sentence structures in the listening passage were complicated for me.	L-S	30	1 (3.3%)	5 (16.7%)	17 (56.7%)	5 (16.7%)	2 (6.7%)
	L-W	30	0	12 (40.0%)	8 (26.7%)	7 (23.3%)	3 (10.0%)
8. There were a lot of important ideas to be processed during the listening passage.	L-S	30	0	1 (3.3%)	8 (26.7%)	13 (43.3%)	8 (26.7%)
	L-W	30	0	2 (6.7%)	5 (16.7%)	19 (63.3%)	4 (13.3%)
9. Important ideas in the listening passage were paraphrased or repeated more than once.	L-S	30	0	12 (40.0%)	4 (13.3%)	10 (33.3%)	4 (13.3%)
	L-W	30	0	7 (23.3%)	9 (30.0%)	13 (43.3%)	1 (3.3%)
10. Ideas in the listening passage were clearly connected.	L-S	30	0	5 (16.7%)	13 (43.3%)	11 (36.7%)	1 (3.3%)
	L-W	30	0	4 (13.3%)	4 (13.3%)	22 (73.3%)	0

Table 8.5: Perceptions of code complexity (*Corruption*)

Code complexity (*Hans Krebs*)	Task type	N	Frequency (percentage)				
			Strongly disagree	disagree	Neutral	agree	Strongly disagree
6. Vocabulary in the listening passage was difficult for me.	L-S	30	1 (3.3%)	5 (16.7%)	10 (33.3%)	11 (36.7%)	3 (10.0%)
	L-W		0	2 (6.7%)	14 (46.7%)	10 (33.3%)	4 (13.3%)
7. Sentence structures in the listening passage were complicated for me.	L-S	30	1 (3.3%)	9 (30.0%)	8 (26.7%)	11 (36.7%)	1 (3.3%)
	L-W		0	4 (13.3%)	16 (53.3%)	7 (23.3%)	3 (10.0%)
8. There were a lot of important ideas to be processed during the listening passage.	L-S	30	0	4 (13.3%)	5 (16.7%)	14 (46.7%)	7 (23.3%)
	L-W		0	2 (6.7%)	5 (16.7%)	17 (56.7%)	6 (20.0%)
9. Important ideas in the listening passage were paraphrased or repeated more than once.	L-S	30	1 (3.3%)	14 (46.7%)	6 (20.0%)	9 (30.0%)	0
	L-W		1 (3.3%)	9 (30.0%)	14 (46.7%)	6 (20.0%)	0
10. Ideas in the listening passage were clearly connected.	L-S	30	0	5 (16.7%)	12 (40.0%)	12 (40.0%)	1 (3.3%)
	L-W		0	3 (10.0%)	14 (46.7%)	13 (33.3%)	3 10.0%)

Table 8.6: Perceptions of code complexity (*Hans Krebs*)

Code complexity (*Talent*)	Task type	N	Frequency (percentage)				
			Strongly disagree	disagree	Neutral	agree	Strongly disagree
6. Vocabulary in the listening passage was difficult for me.	L-S	30	0	12 (40.0%)	11 (36.7%)	6 (20.0%)	1 (3.3%)
	L-W		6 (20.0%)	10 (33.3%)	9 (30.0%)	5 (16.7%)	0
7. Sentence structures in the listening passage were complicated for me.	L-S	30	0	11 (36.7%)	11 (36.7%)	8 (26.7%)	0
	L-W		3 (10.0%)	10 (33.3%)	10 (33.3%)	7 (23.3%)	0
8. There were a lot of important ideas to	L-S	30	0	2	9	18	1

be processed during the listening passage.				(6.7%)	(30.0%)	(60.0%)	(3.3%)
	L-W		0	4 (13.3%)	9 (30.0%)	15 (50.0%)	2 (6.7%)
9. Important ideas in the listening passage were paraphrased or repeated more than once.	L-S	30	0	8 (26.7%)	7 (23.3%)	15 (50.0%)	0
	L-W		0	3 (10.0%)	9 (30.0%)	15 (50.0%)	3 (10.0%)
10. Ideas in the listening passage were clearly connected.	L-S	30	0	1 (3.3%)	12 (40.0%)	14 (46.7%)	3 (10.0%)
	L-W		0	1 (3.3%)	10 (33.3%)	19 (63.3%)	0

Table 8.7: Perceptions of code complexity (*Talent*)

Code complexity (*Vitamin D*)	Task type	N	Frequency (percentage) Strongly disagree	disagree	Neutral	agree	Strongly disagree
6. Vocabulary in the listening passage was difficult for me.	L-S	30	3 (10.0%)	15 (50.0%)	8 (26.7%)	4 (13.3%)	0
	L-W		4 (13.3%)	10 (33.3%)	13 (43.3%)	3 (10.0%)	0
7. Sentence structures in the listening passage were complicated for me.	L-S	30	4 (13.3%)	16 (53.3%)	10 (33.3%)	0	0
	L-W		3 (10.0%)	14 (46.7%)	4 (13.3%)	9 (30.0%)	0
8. There were a lot of important ideas to be processed during the listening passage.	L-S	30	1 (3.3%)	11 (36.7%)	9 (30.0%)	8 (26.7%)	1 (3.3%)
	L-W		1 (3.3%)	1 (3.3%)	16 (50.3%)	7 (23.3%)	5 (16.7%)
9. Important ideas in the listening passage were paraphrased or repeated more than once.	L-S	30	0	5 (16.7%)	8 (26.7%)	14 (46.7%)	3 (10.0%)
	L-W		0	8 (26.7%)	7 (23.3%)	13 (43.3%)	2 (6.7%)
10. Ideas in the listening passage were clearly connected.	L-S	30	0	0	4 (13.3%)	21 (70.0%)	5 (16.7%)
	L-W		0	2 (6.7%)	9 (30.0%)	19 (63.3%)	0

Table 8.8: Perceptions of code complexity (*Vitamin D*)

With respect to code complexity, it can be summarized that *Hans Krebs* is likely to be the most linguistically complex as a majority of the participants considered the textual information to be dense. About half of them thought it was lexically and structurally complex and the important ideas were not paraphrased or repeated. *Talent* and *Vitamin D* were considered less lexically and syntactically complex than *Corruption* and *Hans Krebs*. Although about half of the participants (strongly) agreed on the occurrence of textual redundancy and a clear organization in *Corruption*, about half of them held neutral views on its lexical and structural complexity. Hence, it is likely that *Hans Krebs*, as perceived by this group of participants, was linguistically the most difficult text whereas *Corruption* was neutral in terms of code complexity and *Talent* and *Vitamin D* were comparatively easier.

8.6.2 Cognitive complexity

The investigation of cognitive complexity, as indicated earlier in this section, related to two aspects, i.e. cognitive familiarity and cognitive processing demands.

Cognitive familiarity

Cognitive familiarity is associated with the participants' familiarity with the content (Item 11), the predictability of the content (Item 12), their familiarity with the discourse genre (Item 13), and familiarity with the task requirements (Item 14).

The perceptions of cognitive familiarity for the four listening passages were similar in pattern (see Tables 8.9-8.12). About half of the participants indicated that they were not familiar with the text content (Item 11) or did not think that they could predict the content of the text after listening to the first few sentences (Item 12) – with the highest number of the participants found to share this view for *Hans Krebs* (60.0% for the listening-to-speak task and 80.0% for the listening-to-write task). Almost half of the participants held a neutral view (neither agreed nor disagreed) on familiarity with the task structure (Items 13 and 14).

Cognitive familiarity (*Corruption*)	Task type	N	Frequency (percentage)				
			Strongly disagree	disagree	Neutral	agree	Strongly disagree
11. I am familiar with the content of the listening passage.	L-S	30	5 (16.7%)	10 (33.3%)	11 (36.7%)	4 (13.3%)	0
	L-W		3 (10.0%)	13 (43.3%)	9 (30.0%)	5 (16.7%)	0
12. I could predict the rest of listening content after listening to the first few sentences.	L-S	30	4 (13.3%)	7 (23.3%)	12 (40.0%)	6 (20.0%)	1 (3.3%)
	L-W		5 (16.7%)	15 (50.0%)	7 (23.3%)	3 (10.0%)	0
13. I am familiar with academic	L-S	30	1	2	15	11	1

lectures in English.			(3.3%)	(6.7%)	(50.0%)	(36.7%)	(3.3%)
	L-W		0	2 (6.7%)	16 (53.3%)	9 (30.0%)	3 (10.0%)
14. I am familiar with listening to an academic text and then orally summarizing it/writing its summary.	L-S	30	3 (10.0%)	5 (16.7%)	11 (36.7%)	7 (23.3%)	4 (13.3%)
	L-W		3 (10.0%)	7 (23.3%)	11 (36.7%)	9 (30.0%)	0

Table 8.9: Perceptions of cognitive familiarity (*Corruption*)

Cognitive familiarity (*Hans Krebs*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	disagree	Neutral	agree	Strongly disagree
11. I am familiar with the content of the listening passage.	L-S	30	10 (33.3%)	9 (30.0%)	6 (20.0%)	4 (13.3%)	1 (3.3%)
	L-W		13 (43.3%)	10 (33.3%)	6 (20.0%)	1 (3.3%)	0
12. I could predict the rest of listening content after listening to the first few sentences.	L-S	30	10 (33.3%)	8 (26.7%)	6 (20.0%)	6 (20.0%)	0
	L-W	29*	13 (43.3%)	11 (36.7%)	4 (13.3%)	1 (3.3%)	0
13. I am familiar with academic lectures in English.	L-S	29*	0	0	16 (53.3%)	13 (43.3%)	0
	L-W		1 (3.3%)	3 (10.0%)	16 (53.3%)	9 (30.0%)	1 (3.3%)
14. I am familiar with listening to an academic text and then orally summarizing it/writing its summary.	L-S	30	5 (16.7%)	5 (16.7%)	10 (33.3%)	6 (20.0%)	4 (13.3%)
	L-W	29*	2 (6.7%)	3 (10.0%)	13 (43.3%)	7 (23.3%)	4 (13.3%)

**1 missing*

Table 8.10: Perceptions of cognitive familiarity (*Hans Krebs*)

Cognitive familiarity (*Talent*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	disagree	Neutral	agree	Strongly disagree
11. I am familiar with the content of the listening passage.	L-S	30	1 (3.3%)	13 (43.3%)	5 (16.7%)	2 (36.7%)	0
	L-W		2 (6.7%)	13 (43.3%)	7 (23.3%)	6 (20.0%)	2 (6.7%)
12. I could predict the rest of listening content after listening to the first few sentences.	L-S	29*	5 (16.7%)	10 (33.3%)	7 (23.3%)	6 (20.0%)	1 (3.3%)
	L-W	29*	1 (3.3%)	12 (40.0%)	7 (23.3%)	6 (20.0%)	3 (10.0%)
13. I am familiar with academic lectures in English.	L-S	29*	2 (6.7%)	2 (6.7%)	15 (50.0%)	9 (30.0%)	1 (3.3%)
	L-W	30	0	3 (10.0%)	17 (56.7%)	8 (26.7%)	2 (6.7%)
14. I am familiar with listening to an academic text and then orally summarizing it/writing its summary.	L-S	30	3 (10.0%)	7 (23.3%)	12 (40.0)	2 (6.7%)	6 (20.0%)
	L-W		1 (3.3%)	3 (10.0%)	12 (40.0%)	4 (13.3%)	10 (33.3%)

**1 missing*

Table 8.11: Perceptions of cognitive familiarity (*Talent*)

Cognitive familiarity (*Vitamin D*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	disagree	Neutral	agree	Strongly disagree
11. I am familiar with the content of the listening passage.	L-S	30	4 (13.3%)	9 (30.0%)	9 (30.0%	6 (20.0%)	2 (6.7%)
	L-W		4 (13.3%)	11 (36.7)	9 (30.0%)	4 (13.3%)	2 (6.7%)

12. I could predict the rest of listening content after listening to the first few sentences.	L-S	29*	3 (10.0%)	3 (10.0%)	15 (50.0%)	7 (23.3%)	1 (3.3%)
	L-W		5 (16.7%)	9 (30.0%)	6 (20.0%)	10 (33.3%)	0
13. I am familiar with academic lectures in English.	L-S	30	0	3 (10.0%)	14 (46.7%)	12 (40.0%)	1 (3.3%)
	L-W	29*	0	1 (3.3%)	15 (50.0%)	11 (36.7%)	2 (6.7%)
14. I am familiar with listening to an academic text and then orally summarizing it/writing its summary.	L-S	30	5 (16.7%)	2 (6.7%)	9 (30.0%)	12 (40.0%	2 (6.7%)
	L-W	29*	0	3 (10.0%)	14 (46.7%)	5 (16.7%)	7 (23.3%)

** 1 missing*

Table 8.12: Perceptions of cognitive familiarity (*Vitamin D*)

Thus, in terms of cognitive familiarity, it can be summarized that overall the participants were not familiar with the content of the texts. Almost half of them, in addition, neither agreed nor disagreed that they were familiar with the task structure.

Cognitive processing demands

Cognitive processing demands are associated with five issues, i.e. text organization (Item 15), amount of online computation (Item 16), sufficiency of information (Item 17), clarity of information (Item 18), and type of ideas: abstract or concrete (Item 19). The analysis showed that between 56.6% and 77.6% of the participants perceived all texts as cognitively demanding in terms of amount of online computation. That is, they (strongly) agreed they had to pay attention to more than one idea at a time (Item 16). However, *Vitamin D* and *Talent* were thought to be less cognitively demanding than their counterparts as more than half of the participants (strongly) agreed that the texts had a clear organization (Item 15) and provided sufficient information for task performance (Item 17). Moreover, close to two-thirds of the participants were neutral or did not think that the passages contained a lot of implied meanings (Item 18) or abstract ideas (Item 19). For *Hans Krebs* and *Corruption*, the perceptions were quite similar. The highest number of the participants (between 40.0%-

70.0%) appeared to hold a neutral view on the clarity of the text's organization. About half of them held a neutral view on directness of information for *Corruption*, neither agreeing nor disagreeing that the text's ideas were explicitly indicated.

Cognitive processing demands (*Corruption*)	**Task type**	**N**	**Frequency (percentage)**				
			Strongly disagree	**disagree**	**Neutral**	**agree**	**Strongly disagree**
15. The ideas in the listening passage were organized clearly.	L-S	30	2 (6.7%)	3 (10.0%)	21 (70.0%)	3 (10.0%)	1 (3.3%)
	L-W	29*	1 (3.3%)	4 (13.3%)	12 (40.0%)	10 (33.3%)	2 (6.7%)
16. I had to pay attention to more than one idea at a time.	L-S	30	0	2 (6.7%)	4 (13.3%)	20 (66.7%)	4 (13.3%)
	L-W		0	0	10 (33.3%)	16 (53.3%)	4 (13.3%)
17. The listening passage provided sufficient ideas for me to complete an oral summary.	L-S	30	0	6 (20.0%)	6 (20.0%)	16 (53.3%)	2 (6.7%)
	L-W	29*	0	1 (3.3%)	13 (43.3%)	14 (46.7%)	1 (3.3%)
18. The listening passage contained a lot of implied meanings.	L-S	30	0	2 (6.0%)	16 (53.3%)	6 (20.0%)	2 (6.7%)
	L-W		2 (6.7%)	8 (26.7%)	10 (33.3%)	9 (30.0%)	1 (3.3%)
19. The listening passage contained a lot of abstract ideas	L-S	30	3 (10.0%)	11 (36.7%)	8 (26.7%)	8 (26.7%)	0
	L-W		3 (10.0%)	11 (36.7%)	9 (30.0%)	7 (23.3%)	0

* *1 missing*

Table 8.13: Perceptions of cognitive processing demand (*Corruption*)

Cognitive processing demand (*Hans Krebs*)	**Task type**	**N**	**Frequency (percentage)**				
			Strongly disagree	**disagree**	**Neutral**	**agree**	**Strongly disagree**
15. The ideas in the listening passage were organized clearly.	L-S	30	1 (3.3%)	7 (23.3%)	13 (43.3%)	9 (30.0%)	0
	L-W		3 (10.0%)	6 (20.0%)	18 (60.0%)	2 (6.7%)	1 (3.3%)
16. I had to pay attention to more than one idea at a time.	L-S	30	0	0	10 (33.3%)	13 (43.3%)	7 (23.3%)
	L-W		0	0	13 (43.3%)	14 (46.7%)	3 (10.0%)
17. The listening passage provided sufficient ideas for me to complete an oral summary.	L-S	30	0	8 (26.7%)	9 (30.0%)	12 (40.0%)	1 (3.3%)
	L-W	29*	1 (3.3%)	7 (23.3%)	9 (30.0%)	10 (33.3%)	3 (10.0%)
18. The listening passage contained a lot of implied meanings.	L-S	30	1 (3.3%)	12 (40.0%)	9 (30.0%)	8 (26.7%)	0
	L-W		1 (3.3%)	7 (23.3%)	11 (36.7%)	8 (26.7%)	3 (10.0%)
19. The listening passage contained a lot of abstract ideas	L-S	30	0	6 (20.0%)	12 (40.0%)	10 (33.3%)	2 (6.7%)
	L-W		1 (3.3%)	5 (16.7%)	12 (40.0%)	8 (26.7%)	4 (13.3%)

*1 missing

Table 8.14: Perceptions of cognitive processing demand (*Hans Krebs*)

Cognitive processing demand (*Talent*)	Task type	N	Frequency (percentage)				
			Strongly disagree	**disagree**	**Neutral**	**agree**	**Strongly disagree**
15. The ideas in the listening passage were organized clearly.	L-S	30	0	2 (6.7%)	11 (36.7%)	13 (43.3%)	4 (13.3%)
	L-W	29*	0	2 (6.7%)	9 (30.0%)	17 (56.7%)	2 (6.7%)
16. I had to pay attention to more than one idea at a time.	L-S	30	0	1 (3.3%)	6 (20.0%)	20 (66.7%)	3 (10.0%)
	L-W		0	3 (10.0%)	6 (20.0%)	17 (56.7%)	4 (13.3%)
17. The listening passage provided sufficient ideas for me to complete an oral summary.	L-S	29*	0	1 (3.3%)	9 (30.0%)	14 (46.7%)	5 (16.7%)
	L-W	29*	1 (3.3%)	6 (20.0%)	9 (30.0%)	11 (36.7%)	3 (10.0%)
18. The listening passage contained a lot of implied meanings.	L-S	30	1 (3.3%)	12 (40.0%)	8 (26.7%)	9 (30.0%)	0
	L-W		1 (3.3%)	11 (36.7%)	14 (46.7%)	4 (13.3%)	0
19. The listening passage contained a lot of abstract ideas	L-S	30	0	5 (16.7%)	14 (47.6%)	8 (26.7%)	3 (10.0%)
	L-W		1 (3.3%)	12 (40.0%)	5 (16.7%)	6 (20.0%)	6 (20.0%)

*1 missing

Table 8.15: Perceptions of cognitive processing demand (*Talent*)

Cognitive processing demand (*Vitamin D*)	Task type	N	Frequency (percentage)				
			Strongly disagree	**disagree**	**Neutral**	**agree**	**Strongly disagree**
15. The ideas in the listening passage were organized clearly.	L-S	30	0	2 (6.7%)	6 (20.0%	18 (60.0%)	4 (13.3%)
	L-W		0	0	14 (46.7%)	15 (50.0%)	1 (3.3%)
16. I had to pay attention to more than one idea at a time.	L-S	30	0	3 (10.0%)	8 (26.7%)	16 (53.3%)	3 (10.0%)
	L-W		0	3 (10.0%)	10 (33.3%)	13 (43.3%)	4 (13.3%)
17. The listening passage provided sufficient ideas for me to complete an oral summary.	L-S	30	0	2 (6.7%)	11 (36.7%)	15 (50.0%)	2 (6.7%)
	L-W		0	4 (23.3%)	9 (30.0%)	14 (46.7%)	3 (10.0%)
18. The listening passage contained a lot of implied meanings.	L-S	30	5 (16.7%)	18 (60.0)	6 (20.0%)	1 (3.3%)	0
	L-W		4 (13.3%)	9 (30.0)	10 (33.3%)	7 (23.3%)	0
19. The listening passage contained a lot of abstract ideas	L-S	30	9 (30.0%)	14 (46.7%)	4 (13.3%)	3 (10.0%)	0
	L-W		3 (10.0%)	13 (43.3%)	8 (26.7%)	6 (20.0%)	0

*1 missing

Table 8.16: Perceptions of cognitive processing demand (*Vitamin D*)

Based on the results, it can be concluded that *Hans Krebs* and *Corruption* seem to be more cognitively demanding than *Vitamin D* and *Talent* because of unclear text organization and because more than one idea needed to be processed at a time, according to the participants. *Vitamin D* and *Talent*, on the other hand, were considered to provide sufficient ideas for task completion and these ideas were not implicit or abstract.

With regard to cognitive complexity, the analysis showed that most of the participants were not familiar with the content of the listening passages or did not feel that they were able to predict the rest of the text correctly after listening to the first few sentences. The listening tasks were found cognitively demanding, especially in terms of online computation. The tasks, as perceived by the participants, contained several key ideas that they had to pay attention to simultaneously. Taking into account all factors contributing to cognitive complexity, it seems that they found *Hans Krebs* the most cognitively demanding task of the four, whereas *Vitamin D* appeared to be the least cognitively complicated task in their view.

8.6.3 Communication Stress

Communication stress involves four areas of task characteristics that could possibly cause stress in task communication. They are time pressure, time limit, length of text, and speed of presentation. The analysis showed that almost half of the participants felt under time pressure (Item 20) while performing the listening-to-speak tasks in all listening passages (60.0% for *Corruption*, *Hans Krebs*, and *Vitamin D*, and 46.7% for *Talent*). This number was found to be two times higher than those indicating feeling under time pressure in the listening-to-write tasks (36.7% for *Corruption*, 33.3% for *Hans Krebs*, 23.4% for *Talent* and 20.0% for *Vitamin D*).

With regard to the time limit (Item 21), almost half of the participants (46.7%) did not find that they had enough time to produce an oral summary in the listening-to-speak task with *Corruption*, whereas for the other listening passages, they held a neutral view on the time allowed. A large number of the participants, however, agreed that they had enough time to perform the listening-to-write tasks (60.0% for *Corruption*, 56.7% for *Hans Krebs*, 80.0% for *Talent*, and 63.4% for *Vitamin D*).

In terms of text length (Item 22), about half of the participants did not agree that the texts were too long while performing both the listening-to-speak and the listening-to-write tasks, except for the *Corruption* text, where slightly over half of the participants (63.3% in the listening-to-speak and 53.3% in the listening-to-write tasks) remained neutral about the text length, neither agreeing or disagreeing whether the text was too long.

Close to 50% of the participants held a neutral view on the speed of text presentation (Item 23), neither agreeing nor disagreeing that the text was spoken too fast when listening to *Corruption*, *Hans Krebs*, and *Talent* in both the listening-to-speak and listening-to-write tasks. The exception is for *Vitamin D*, where about half of the participants clearly indicated that the text was not too fast (56.7% for the listening-to-speak and 46.6% for the listening-to-write tasks).

Communication stress (*Corruption*)	**Task type**	**N**	**Frequency (percentage)**				
			Strongly disagree	**disagree**	**Neutral**	**Agree**	**Strongly disagree**
20. I felt under time pressure while performing the task.	L-S	30	1 (3.3%)	4 (13.3%)	7 (23.3%)	14 (46.7%)	4 (13.3%)
	L-W		2 (6.7%)	6 (20.0%)	11 (36.7%)	9 (30.0%)	2 (6.7%)
21. I had enough time to perform the task.	L-S	30	1 (3.3%)	4 (13.3%)	11 (36.7%)	12 (40.0%)	2 (6.7%)
	L-W		6 (20.0%)	12 (40.0%)	10 (33.3%)	2 (6.7%)	0
22. The listening was too long.	L-S	30	4 (13.3%)	4 (13.3%)	19 (63.3%)	3 (10.0%)	0
	L-W		4 (13.3%)	9 (30.0%)	16 (53.3%)	1 (3.3%)	0
23. The passage was spoken too fast.	L-S	30	1 (3.3%)	7 (23.3%)	16 (53.3%)	6 (20.0%)	0

	L-W		4 (13.3%)	7 (23.3%)	12 (40.0%)	6 (20.0%)	1 (3.3%)

Table 8.17: Perceptions of communication stress (*Corruption*)

Communication stress (*Hans Krebs*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	disagree	Neutral	Agree	Strongly disagree
20. I felt under time pressure while performing the task.	L-S	30	0	3 (10.0%)	9 (30.0%)	13 (43.3%)	5 (16.7%)
	L-W		4 (13.3%)	3 (10.0%)	13 (43.3%)	7 (23.3%)	3 (10.0%)
21. I had enough time to perform the task.	L-S	30	2 (6.7%)	11 (36.7%)	11 (36.7%)	5 (16.7%)	1 (3.3%)
	L-W		11 (36.7%)	6 (20.0%)	7 (23.3%)	6 (20.0%)	0
22. The listening was too long.	L-S	30	0	15 (50.0%)	13 (43.3%)	2 (6.7%)	0
	L-W		8 (26.7%)	9 (30.0%)	9 (30.0%)	3 (10.0%)	1 (3.3%)
23. The passage was spoken too fast.	L-S	30	1 (3.3%)	10 (33.3%)	12 (40.0%)	6 (20.0%)	1 (3.3%)
	L-W		0	5 (16.7%)	16 (53.3%)	7 (23.3%)	2 (6.7%)

Table 8.18: Perceptions of communication stress (*Hans Krebs*)

Communication stress (*Talent*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	Disagree	Neutral	agree	Strongly disagree
20. I felt under time pressure while performing the task.	L-S	30	0	6 (20.0%)	10 (33.3%)	12 (40.0%)	2 (6.7%)
	L-W		7 (23.3%)	8 (26.7%)	8 (26.7%)	5 (16.7%)	2 (6.7%)
21. I had enough time to perform the task.	L-S	30	2 (6.7%)	7 (23.3%)	11 (36.7%)	8 (26.7%)	2 (6.7%)
	L-W		11 (36.7%)	13 (43.3%)	4 (13.3%)	0	2 (6.7%)
22. The listening was too long.	L-S	29*	2 (6.7%)	11 (36.7%)	13 (43.3%)	3 (10.0%)	0
	L-W	30	9 (30.0%)	13 (43.3%)	7 (23.3%)	1 (3.3%)	0
23. The passage was spoken too fast.	L-S	30	2 (6.7%)	6 (20.0%)	14 (46.7%)	8 (26.7%)	0
	L-W		3 (10.0%)	8 (26.7%)	14 (46.7%)	4 (13.3%)	1 (3.3%)

**1 missing*

Table 8.19: Perceptions of communication stress (*Talent*)

Communication stress (*Vitamin D*)	**Task type**	**N**	**Frequency (percentage)** Strongly disagree	Disagree	neutral	agree	Strongly disagree
20. I felt under time pressure while performing the task.	L-S	30	1 (3.3%)	5 (16.7%)	4 (13.3%)	16 (53.3%)	4 (13.3%)
	L-W		2 (6.7%)	7 (23.3%)	15 (50.0%)	5 (16.7%)	1 (3.3%)
21. I had enough time to perform the task.	L-S	30	5 (16.7%)	9 (30.0%)	9 (30.0%)	6 (20.0%)	1 (3.3%)
	L-W		8 (26.7%)	11 (36.7%)	7 (23.3%)	4 (13.3%)	0
22. The listening was too long.	L-S	30	8 (26.7%)	7 (23.3%)	15 (50.0%)	0	0
	L-W		4	13	11	2	0

			(13.3%)	(43.3%)	(36.7%)	(6.75)	
23. The passage was spoken too fast.	L-S	30	5 (16.7%)	12 (40.0%)	8 (26.7%)	5 (16.7%)	0
	L-W		1 (3.3%)	13 (43.3%)	11 (36.7)	5 (16.7%)	0

Table 8.20: Perceptions of communication stress (*Vitamin D*)

To sum up, the listening-to-speak tasks were likely to be more stressful than the listening-to-write tasks as more participants reported being stressed while performing the listening-to-speak tasks in all four listening topics than in the listening-to-write tasks. The stress was, however, the result of the time pressure imposed by the task rather than the length and the speed of delivery of the listening texts.

Overall, task difficulty, as perceived by the participants, varied. Many participants agreed that the textual information of all the listening passages was dense. *Corruption* and *Hans Krebs* in particular were not clearly organized and most of the participants indicated that they were not familiar with the ideas in these two texts. Hence, *Hans Krebs* and *Corruption*, according to the participants, are likely to be more linguistically and cognitively complex than *Vitamin D* and *Talent*. In addition, most of the participants reported being under time pressure especially while performing the listening-to-summarize tasks.

8.7 Relationship between perceptions of listening difficulty and task performance

This section provides the findings on the relationship between the participants' perceptions of listening difficulty and their performance on the tasks. Adhering to the PTE Academic scoring criteria, the listening-to-speak task performances were scored on content, pronunciation, and fluency and the listening-to-write tasks were scored on content, grammar, and vocabulary. A Spearman[8] correlation was then run in order to investigate the relationship between the perceptions of listening difficulty and task performance. Tables 8.21-8.28 list the correlation results, according to the listening passages and tasks investigated.

An examination of the relationship between the perceptions of listening difficulty and task performance overall pointed to a few significantly strong correlations between perceptions of task difficulty and performance when taking the correlation cut-off points indicated in Field (2005), that values of ± 0.1 represent a small effect, ±0.3 is a medium effect, and ±0.5 is a large effect. Although weak and moderate correlations in both positive and negative directions were more common, considerably more correlation values were not significant.

For *Corruption*, more significant values were found in the listening-to-speak task than in the listening-to-write task. In the listening-to-speak tasks, the content scores were negatively related to the perceptions of lexical difficulty (r=-.415*) and syntactic complexity (r=-.580**), meaning that the participants who found the texts lexically and syntactically difficult provided less accurate summary content. The content score was positively related to the perceptions of content predictability (r=3.68*) and information organization (r=.562**). This means that the participants who indicated being able to predict the text coming next in their listening and recognized the texts' ideas as being

---

[8] A Spearman correlation is a correlation analysis used when data are not normally distributed. In this study, the statistical analyses showed that the degrees of skewness and kurtosis were greater than ±2 in all the data set, suggesting that the data were not normally distributed.

clearly organized were likely to be better in providing oral summary content. Having a sufficiency of information for task performance was related to pronunciation (r=.370*), showing that the participants who agreed that they had enough information to orally summarize the text were likely to score highly on pronunciation. Fluency was found to be negatively related to the perceptions of syntactic complexity (r=-.512**) and clarity of information (r=-.383*), indicating that the participants who perceived the text as being structurally complex and containing a lot of implied meaning were likely to score low on fluency.

In the *Corruption* listening-to-write task, only two significant values were found. That is the positive correlation between the perceptions of time pressure and content score (r=.369*) and vocabulary score (r=.407*), meaning that the participants who felt under time pressure while listening were likely to provide better content and more relevant vocabulary in their written summary.

No significant relationship between the perceptions of communication stress and the performance on the listening-to-speak task was found. In the listening-to-write tasks, only the perceptions of communication stress (time pressure) were found to be related to task performance, but not perceptions of code complexity nor those of cognitive complexity.

***Corruption*: Listening-to-speak task**	**Scores**		
	Cont.	**Pron.**	**Flu.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	-.415*	.209	-.130
7. Sentence structures in the listening passage were complicated for me.	-.508**	-.135	-.512**
8. There were a lot of important ideas to be processed during the listening passage.	-.313	.035	-.109
9. Important ideas in the listening passage were paraphrased or repeated more than once.	-.237	-.089	-.290
10. Ideas in the listening passage were clearly connected.	.263	.225	.179
Cognitive complexity			

11. I am familiar with the content of the listening passage.	.310	.158	.212
12. I could predict the rest of listening content after listening to the first few sentences.	.368*	.221	.251
13. I am familiar with academic lectures in English.	.174	.054	.165
14. I am familiar with listening to an academic text and then orally summarizing it.	.169	-.325	-.179
15. The ideas in the listening passage were organized clearly.	.562**	.103	.200
16. I had to pay attention to more than one idea at a time.	-.238	.075	-.031
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.298	.370*	.320
18. The listening passage contained a lot of implied meanings.	-.303	-.067	-.383*
19. The listening passage contained a lot of abstract ideas	-.306	-.047	-.160
Communication stress			
20. I felt under time pressure while performing the task.	-.288	-.218	-.244
21. I had enough time to perform the task.	-.254	-.284	-.330
22. The listening was too long.	-.277	.030	-.132
23. The passage was spoken too fast.	-.287	-.193	-.196

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.21: The relationship between the perceptions of task difficulty and task performance (*Corruption*: Listening-to-speak task)

***Corruption*: Listening-to-write task**	**Scores**		
	Cont.	**Gram.**	**Vocab.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	.040	-.068	.154
7. Sentence structures in the listening passage were complicated for me.	-.095	-.038	-.087
8. There were a lot of important ideas to be processed during the listening passage.	.082	-.079	.177
9. Important ideas in the listening passage were paraphrased or repeated more than once.	-.454	-.025	-.161
10. Ideas in the listening passage were clearly connected.	-.237	-.267	-.263
Cognitive complexity			
11. I am familiar with the content of the listening passage.	-.177	-.072	-.306
12. I could predict the rest of listening content after listening to the first few sentences.	-.197	-.242	-.220
13. I am familiar with academic lectures in English.	-.164	.427	.123
14. I am familiar with listening to an academic text and then orally summarizing it.	.355	.151	.230
15. The ideas in the listening passage were organized clearly.	.226	.221	-.089
16. I had to pay attention to more than one idea at a time.	.055	-.260	-.064
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.079	.195	.029
18. The listening passage contained a lot of implied meanings.	-.331	-.063	-.016
19. The listening passage contained a lot of abstract ideas	-.128	.194	.166
Communication stress			
20. I felt under time pressure while performing the task.	.369*	.232	.407*
21. I had enough time to perform the task.	.085	.226	.329
22. The listening was too long.	-.110	.148	.150
23. The passage was spoken too fast.	-.165	.067	-.209

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.22: The relationship between the perceptions of task difficulty and task performance (*Corruption*: Listening-to-write tasks)

Overall for *Corruption*, more correlations between task performance and perceptions of task difficulty were found for the listening-to-speak task than the listening-to-write task. In the former task, although more significant values were found, only the correlation between performance and the perceptions of syntactic complexity and of information organization were strong in nature. The participants' perceptions of communication stress were not found to be significantly related to any aspect of task performance in the listening-to-speak task, but significantly related to content and fluency in the listening-to-write task at low and moderate degrees.

For the *Hans Krebs* text (see Tables 8.23 and 8.24), a small number of significant values were found. In the listening-to-speak task, only fluency was found to negatively correlate with the perceptions of discourse familiarity (r=-.404*) and positively with the time limit (r=.381*). The values indicate that the participants who thought that they were familiar with academic lectures were likely to score low in fluency, whereas those considering that they had limited time in task performance were likely to score high on fluency. The values however show a moderate effect only.

In the *Hans Krebs* listening-to-write task, the performance on grammar was negatively related to the perceptions of information density (r=.-.362*) and the speech delivery of the spoken text (r=.-374*), meaning that the participants who agreed that the information in the text was dense and was spoken too fast were likely to score low on grammar. In addition, it was found that the fluency scores were positively related to the perceptions of content familiarity (r=.386*), the time limit (r=.415*), and the speed of the text (r=.390*), indicating that the participants who agreed that they were familiar with the content, had enough time to complete the task, or found that the text was spoken not too fast were likely to score high on fluency. Interestingly, no significant correlations were found, however, between content score and perceptions of code complexity, cognitive complexity, or communication stress in both the listening-to-speak and the listening-to-write tasks.

***Hans Krebs*: Listening-to-speak task**	**Scores**		
	Cont.	**Pron.**	**Flu.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	-.163	-.070	.019
7. Sentence structures in the listening passage were complicated for me.	.177	-.297	.213
8. There were a lot of important ideas to be processed during the listening passage.	.025	-.173	.059
9. Important ideas in the listening passage were paraphrased or repeated more than once.	.128	.015	.162
10. Ideas in the listening passage were clearly connected.	-.008	-.122	-.143
Cognitive complexity			
11. I am familiar with the content of the listening passage.	-.050	-.071	-.228
12. I could predict the rest of listening content after listening to the first few sentences.	-.081	-.109	-.066
13. I am familiar with academic lectures in English.	-.145	-.153	-.404*
14. I am familiar with listening to an academic text and then orally summarizing it.	.260	.089	.175
15. The ideas in the listening passage were organized clearly.	.146	-.083	.137
16. I had to pay attention to more than one idea at a time.	-.024	-.018	-.075
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.102	.195	-.111
18. The listening passage contained a lot of implied meanings.	-.109	.125	.192
19. The listening passage contained a lot of abstract ideas	-.247	.016	-.022
Communication stress			
20. I felt under time pressure while performing the task.	.252	-.235	.171
21. I had enough time to perform the task.	-.094	-.088	.381*
22. The listening was too long.	-.061	-.243	.242
23. The passage was spoken too fast.	-.049	-.046	.232

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.23: The relationship between the perceptions of task difficulty and task performance (*Hans Krebs*: Listening-to-speak task)

Hans Krebs: Listening-to-write tasks	**Scores**		
	Cont.	**Gram.**	**Vocab.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	.100	.203	.304
7. Sentence structures in the listening passage were complicated for me.	.024	-.012	.087
8. There were a lot of important ideas to be processed during the listening passage.	-.051	-.362*	-.256
9. Important ideas in the listening passage were paraphrased or repeated more than once.	.258	.262	.407*
10. Ideas in the listening passage were clearly connected.	.050	-.279	-.241
Cognitive complexity			

11. I am familiar with the content of the listening passage.	.058	.297	.368*
12. I could predict the rest of listening content after listening to the first few sentences.	.183	.156	.299
13. I am familiar with academic lectures in English.	.118	.101	.096
14. am familiar with listening to an academic text and then orally summarizing it.	-.012	-.008	-.118
15. The ideas in the listening passage were organized clearly.	-.047	.084	-.005
16. I had to pay attention to more than one idea at a time.	.155	.000	-.030
17. The listening passage provided sufficient ideas for me to complete an oral summary.	-.009	-.304	-.309
18. The listening passage contained a lot of implied meanings.	-.150	.254	.158
19. The listening passage contained a lot of abstract ideas	-.180	.142	.012
Communication stress			
20. I felt under time pressure while performing the task.	.009	.200	.225
21. I had enough time to perform the task.	.166	.336	.415*
22. The listening was too long.	.000	.142	.239
23. The passage was spoken too fast.	-.279	-.374*	-.390*

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.24: The relationship between the perceptions of task difficulty and task performance (*Hans Krebs*: Listening-to-write task)

Thus, for *Hans Krebs*, a few significant correlations were found for both the listening-to-speak and the listening-to-write tasks, but the content scores were not significantly related to any aspect of the perceptions of task difficulty. In the listening-to-speak task, only fluency was related to one area of cognitive complexity (familiarity with the text discourse) and one of communication stress (time limit). In the listening-to-write task, grammar and vocabulary were significantly related to the perceptions of code complexity and those of communication stress, however with a weak degree of relationship.

For *Talent*, both negative and positive correlations were found with low and moderate effects. In the listening-to-speak task, the content score was found to be positively correlated with syntactic structure (r=.478*), discourse complexity (r=.419*), discourse familiarity (r=.415*), and clarity of information (r=.398*), meaning that the participants who perceived the text as structurally complex, the text's ideas as clearly connected, who were familiar with lectures in English, or thought the text contained a lot of implied meaning were likely to be better at orally summarizing the content of the summary. The content score was in addition negatively related to perceptions of time pressure (r=-.411*) and time limitation (r=-.392*), indicating that the participants who felt under time pressure and time restrictions were likely to score low on content. Pronunciation was positively related to perceptions of sufficiency of information (r=.441*), showing that the participants who agreed that the task provided sufficient information for task completion were likely to score better on pronunciation. Fluency was positively related to perceptions of discourse complexity (r=.416*) and predictability of content (r=.459*), meaning that the participants who thought the ideas in the text were clearly connected and the content of the listening passage predictable were likely to orally summarize the text with more fluency.

In the *Talent* listening-to-write task (see Table 8.25 and 8.26), perceptions of content familiarity were not significantly related to summary content but positively related to grammar, indicating that the participants who thought they were familiar with the content were likely to perform better on the grammar criterion. Perceptions of the familiarity of the task structure were positively related to the vocabulary score (r=.374*), indicating that the participants who thought that they had experienced this type of task before were likely to score higher on vocabulary. Perceptions of the sufficiency of information were positively related to content scores (r=.419*) and vocabulary scores (r=.478*), meaning that the participants who thought the passage provided sufficient ideas for task completion were likely to write a summary with better content and vocabulary.

***Talent*: Listening-to-speak task**	**Scores**		
	Cont.	**Pron.**	**Flu.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	-.169	.030	-.059
7. Sentence structures in the listening passage were complicated for me.	.478**	-.185	.275
8. There were a lot of important ideas to be processed during the listening passage.	-.339	-.171	-.094
9. Important ideas in the listening passage were paraphrased or repeated more than once.	.266	.032	.170
10. Ideas in the listening passage were clearly connected.	.419*	.131	.416*
Cognitive complexity			
11. I am familiar with the content of the listening passage.	.158	.293	.184
12. I could predict the rest of listening content after listening to the first few sentences.	.288	.227	.459*
13. I am familiar with academic lectures in English.	.415*	.014	.159
14. I am familiar with listening to an academic text and then orally summarizing it.	-.149	.037	.114
15. The ideas in the listening passage were organized clearly.	.244	.176	.203
16. I had to pay attention to more than one idea at a time.	.080	.012	-.151
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.257	.441*	.220
18. The listening passage contained a lot of implied meanings.	.398*	-.149	.140
19. The listening passage contained a lot of abstract ideas	-.231	.080	-.025
Communication stress			
20. I felt under time pressure while performing the task.	-.411*	-.181	-.085
21. I had enough time to perform the task.	-.392*	.156	.132
22. The listening was too long.	-.074	.028	.242
23. The passage was spoken too fast.	-.084	-.262	-.367*

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.25: The relationship between the perceptions of task difficulty and task performance (*Talent*: Listening-to-speak task)

***Talent*: Listening-to-write task**	**Scores**		
	Cont.	**Gram.**	**Vocab.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	-.231	-.075	-.335
7. Sentence structures in the listening passage were complicated for me.	-.258	-.111	-.444
8. There were a lot of important ideas to be processed during the listening passage.	-.192	-.323	-.131
9. Important ideas in the listening passage were paraphrased or repeated more than once.	-.325	-.228	-.194
10. Ideas in the listening passage were clearly connected.	-.134	-.094	.357
Cognitive complexity			
11. I am familiar with the content of the listening passage.	.183	.393*	.280
12. I could predict the rest of listening content after listening to the first few sentences.	.053	-.265	.203
13. I am familiar with academic lectures in English.	-.261	.040	.283
14. I am familiar with listening to an academic text and then orally summarizing it.	-.075	-.194	.374*
15. The ideas in the listening passage were organized clearly.	-.096	.209	-.010
16. I had to pay attention to more than one idea at a time.	-.061	-.283	-.069
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.419*	-.035	.478**
18. The listening passage contained a lot of implied meanings.	-.411*	-.318	-.182

19. The listening passage contained a lot of abstract ideas	.473**	.120	.309
Communication stress			
20. I felt under time pressure while performing the task.	.036	-.318	.276
21. I had enough time to perform the task.	-.124	-.190	-.276
22. The listening was too long.	.076	.239	.000
23. The passage was spoken too fast.	.147	-.037	.015

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.26: The relationship between the perceptions of task difficulty and task performance (*Talent*: Listening-to-write task)

For *Talent*, more significant correlations between performance and perceptions were found in the listening-to-speak task than in the listening-to-write task. Content scores were moderately or to a small extent related to all aspects of task difficulty (code complexity, cognitive complexity, and communication stress). Pronunciation only related to one aspect of cognitive complexity (sufficiency of information) and fluency related to only one area of code complexity (discourse complexity) and one aspect of cognitive complexity (predictability of content). For the listening-to-write task, no correlations were found between task performance and perceptions of task difficulty and communication stress. A few significant correlations were found between the perceptions of cognitive complexity and grammar and vocabulary scores, but with a small effect size. Content scores were however found to moderately relate to two areas of cognitive complexity (the perceptions of sufficiency of information and the abstractness of information).

In *Vitamin D*, where most of the participants scored high on content, it was interestingly found that the perceptions of the text’s discourse complexity were negatively related to content scores (r=-.423*), pronunciation (r=-.534*), and fluency (r=-.429*) in the listening-to-speak task. In other words, the participants who agreed that the ideas in the passage were clearly connected were likely to provide an oral summary with less accurate content, less accurate pronunciation and less fluency. Pronunciation was positively related to the time limit (r=.553*), meaning that the participants who agreed they had enough time to perform the task were likely to receive higher scores on pronunciation. The participants in addition were likely to score better on fluency when they agreed that they had experience doing this type of task (r=.424*) and had enough time to perform the task (r=.393*).

In the *Vitamin D* listening-to-write task, the content scores negatively correlated with information type (r=.-495**) and speed of the text (r=-.431*), meaning that the participants agreeing that the text contained a lot of abstract ideas and that it was spoken too fast were likely to score low on content. The participants who thought the ideas in the text were paraphrased or repeated more than once and the text was spoken not too fast were likely to write a summary with more accurate vocabulary (r=.431* and r=-.456*, respectively).

***Vitamin D*: Listening-to-speak task**	**Scores**		
	Cont.	**Pron.**	**Flu.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	-.345	.197	-.008
7. Sentence structures in the listening passage were complicated for me.	-.162	.125	.078
8. There were a lot of important ideas to be processed during the listening passage.	.176	.296	.275
9. Important ideas in the listening passage were paraphrased or repeated more than once.	-.085	-.131	-.251
10. Ideas in the listening passage were clearly connected.	-.423*	-.524**	-.429*
Cognitive complexity			
11. I am familiar with the content of the listening passage.	.302	-.024	-.176
12. I could predict the rest of listening content after listening to the first few sentences.	.243	-.131	.000
13. I am familiar with academic lectures in English.	-.112	.022	-.106
14. I am familiar with listening to an academic text and then orally summarizing it.	.311	.132	.424*
15. The ideas in the listening passage were organized clearly.	-.262	-.186	-.322
16. I had to pay attention to more than one idea at a time.	.164	-.127	.215
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.197	.092	-.257
18. The listening passage contained a lot of implied meanings.	.194	.017	.263
19. The listening passage contained a lot of abstract ideas	-.419*	-.059	.021

Communication stress			
20. I felt under time pressure while performing the task.	-.231	.292	.068
21. I had enough time to perform the task.	.051	.553**	.393*
22. The listening was too long.	.048	.239	.083
23. The passage was spoken too fast.	-.068	.235	-.007

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.27: The relationship between the perceptions of task difficulty and task performance (*Vitamin D*: Listening-to-speak)

Vitamin D: Listening-to-write tasks	**Scores**		
	Cont.	**Gram.**	**Vocab.**
Code complexity			
6. Vocabulary in the listening passage was difficult for me.	.291	.231	.140
7. Sentence structures in the listening passage were complicated for me.	.013	.239	-.056
8. There were a lot of important ideas to be processed during the listening passage.	.098	.272	-.139
9. Important ideas in the listening passage were paraphrased or repeated more than once.	-.053	-.006	.431*
10. Ideas in the listening passage were clearly connected.	.160	.063	-.219
Cognitive complexity			
11. I am familiar with the content of the listening passage.	.068	-.109	.136
12. I could predict the rest of listening content after listening to the first few sentences.	-.067	.056	.065
13. I am familiar with academic lectures in English.	-.218	.002	.168
14. I am familiar with listening to an academic text and then orally summarizing it.	-.152	-.018	.171
15. The ideas in the listening passage were organized clearly.	-.264	-.030	-.028
16. I had to pay attention to more than one idea at a time.	.102	.053	-.099
17. The listening passage provided sufficient ideas for me to complete an oral summary.	.058	.238	.131
18. The listening passage contained a lot of implied meanings.	-.252	.171	.081
19. The listening passage contained a lot of abstract ideas	-.495**	.042	-.146
Communication stress			
20. I felt under time pressure while performing the task.	-.223	.227	-.165
21. I had enough time to perform the task.	-.274	.096	-.176
22. The listening was too long.	-.248	.358	-.109
23. The passage was spoken too fast.	-.431*	-.058	-.456*

*** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level*

Table 8.28: The relationship between the perceptions of task difficulty and task performance (*Vitamin D*: Listening-to-write task)

To conclude, for *Vitamin D*, perceptions of task difficulty were likely to be more related to task performance in the listening-to-speak task than in the listening-to-write task. Perceptions of discourse complexity were negatively correlated with all aspects of performance in the listening-to-speak task. Perceptions of cognitive complexity were not shown to relate to the quality of written summary content, but the perceptions of information type (that the passage contained a lot of abstract ideas) were found to negatively relate to written summary content. Perceptions of the time limit were found to positively relate to pronunciation and fluency, and perceptions on the speediness of the text were shown to negatively correlate with oral summary content and fluency scores.

Overall, the examination of the relationship between perceptions of listening tasks' difficulty and task performances showed both negative and positive relationships between these factors. Significant correlations, however, were found more in the listening-to-speak than in the listening-to-write task. With regard to code complexity, participants perceiving the text as lexically and syntactically complex were likely to orally summarize less accurate content and were less fluent in the task where C*orruption* was used. The clearly connected ideas of the *Vitamin D* text were found to negatively and significantly relate to aspects of task performance (the content summary, pronunciation and fluency), showing that the participants who perceived the text's ideas as being clearly connected were unlikely to perform well on the listening-to-speak tasks. Most aspects of perceptions of code complexity were, however, not significantly related to task performance, especially in *Hans Krebs*, where the perceptions of code complexity were not in any case related to task performance. Perceptions of cognitive complexity were unlikely to be significantly related to the performances on *Hans Krebs* and *Vitamin D*, since most of the correlation values were not significant. Perceptions of information clarity were, on the other hand, found to positively relate to the summary content of *Corruption*. For *Talent*, the participants who perceived the tasks as not cognitively demanding were likely to perform better on the tasks. One aspect of communication which seems to be associated with task performance is time pressure since it was shown that participants feeling under time pressure were likely to perform less well on the tasks with *Talent* and *Corruption*.

Based on the results, it can be concluded that perceptions of listening task difficulty overall are slightly but not strongly associated with task performance. Although a large number of participants perceived *Hans Krebs* and *Corruption* as lexically and cognitively complex texts in which several of the ideas needed to be processed at the same time, these perceptions however do not appear to relate to task performance. Perceptions of the difficulty of *Hans Krebs*, in particular, were not in any respect significantly related to the oral or written summary content. Although perceptions of cognitive complexity and communication stress in the same text were significantly related to fluency in the listening-to-speak task and grammar and vocabulary in the listening-to-write tasks, only a few significant and low correlation values were found. Likewise, although in *Vitamin D* the text was perceived to be less linguistically complex and less cognitively demanding, this was not found to be positively related to the performance. Participants who perceived the text as being clearly organized, in particular, were unlikely to perform well in the listening-to-summarize tasks. For *Corruption*, only lexical and syntactic complexities were found to negatively relate to summary content, and this was found only in the listening-to-summarize task, not in the listening-to-write task. More significant correlations were found in *Talent* than in the other three listening texts. However, since the significant values suggest a weak and moderate correlation and a larger number of correlations were not significant, it can be inferred that perceptions of task difficulty in *Talent* were not strongly related to the task performance.

8.8 Summary

In sum, with respect to perceptions of task and task difficulty, it can be concluded that overall the participants agreed that both the listening-to-speak and the listening-to-write tasks were authentic and a fair way to assess their English academic abilities. The majority of the participants agreed that both task modalities were able to simulate the tasks they encountered in real-life contexts, and the tasks not only required using the abilities needed in their academic studies but were also perceived to be an accurate way to evaluate their academic English, especially their listening skills. In all listening topics, the majority of the participants did not agree that they were familiar with the content and could correctly predict the text based on initial linguistic input. *Hans Krebs* and *Corruption* were perceived to be more linguistically and more cognitively complex than *Talent* and *Vitamin D*, especially in terms of structural and discourse complexity, the amount of ideas that needed to be processed at a time, and the clarity of information. *Vitamin D* and *Talent*, as perceived by the participants, were less lexically and structurally complex and their key ideas were clearly organized and paraphrased or repeated more than once. The participants' perceptions of listening task difficulty were, however, only slightly and weakly related to task performance overall.

Chapter 9 Discussion

9.1 Introduction

This chapter discusses the results in relation to the two aspects of listening the research set out to investigate, i.e. 1) listening comprehension processing in listening-to-summarize tasks, and 2) perceptions of tasks and of listening task difficulty. The results on test-takers' overall cognitive and strategic processing for listening comprehension are first discussed (9.2.1). Next, the role of strategies in listening task performance and the effects of different modalities (speaking and writing) on listening processing are discussed in 9.2.2 and 9.2.3. Section 9.3 discusses the role of listening comprehension in listening-to-summarize tasks and the listening abilities assessed by these tasks. The validity of the listening-to-summarize tasks is then discussed in 9.4, supported by evidence from task processing behaviours and test-takers' perceptions of tasks and of listening task difficulty. Section 9.5 summarizes the discussion.

9.2 The processes and strategies used in comprehending listening input in listening-to-summarize tasks

With the aim of describing the listening abilities assessed by listening-to-summarize tasks, this study set out to investigate the cognitive processes and the strategies test-takers engage in in comprehending the listening input – as reflected in RQ1 and its sub-question. Therefore, this section discusses the listening processes and strategies used overall by the participants, their usefulness and limitations, and other aspects that affected the activation of the processes and strategies.

9.2.1 Overall processes and strategies used to comprehend listening input

The findings presented in Chapters 4-7 revealed that to comprehend the listening input in listening-to-summarize tasks, the participants engaged in a number of processes and strategies. Accurately understanding the input texts and being able to correctly present their main point in the summary tasks required text processing at both lower- and higher-levels. However, the activation of cognitive processes, especially at the higher-level, was unlikely to be successful without the effective use of strategies.

With regard to the cognitive processing, the results showed that complete text understanding occurred at the higher-level of text processing which entails semantic processing at the local and the global levels and pragmatic processing. This understanding however is unlikely to take place without effective processing at the lower-level, which involves acoustic-phonetic processing, word recognition and parsing. It is important to note that in three listening inputs – *Corruption*, *Hans Krebs*, and *Talent*, the participants engaged in both semantic processing and pragmatic processing in order to fully understand the texts, whereas in *Vitamin D* the participants who received a full mark on content did not report engaging in pragmatic processing. Compared to the other listening inputs, the main idea of *Vitamin D* is more explicitly indicated and the text is organized in a way that the participants expected to hear. Therefore, it can be suggested that textual organization and the explicitness of a text's ideas may link to the activation of higher-level processes. The more complex the text's structure and the less explicit the main point, the more likely higher-level processes will be activated.

The successful listeners (who received a full mark on the summary content) were found to process the input texts at the lower level (linguistic processing) more automatically and effectively than did the low and average scorers. Although the high scorers reported difficulties, for example, understanding a technical term such as 'tetany' in *Vitamin D*, they relied on the context to infer its likely meaning 'a kind of disease'. This kind of processing, according to Rost (2011), is approximation – a compensation strategy taking place when listeners superordinate or construct a less precise meaning of problematic words. The successful listeners did not report using any fixation or pausing to think about or figure out the meaning of common words such as 'brain' or 'difficult', while the less successful listeners did. And even when they paused to figure out words, the less successful listeners did not manage to establish the correct words. For instance, the word 'brain' was identified by them as 'bright', and 'difficulty' as 'different'. Rost (2011) calls this 'substitution', a process occurring when listeners substitute unknown words or un-comprehended concepts with known or familiar words or concepts. In this study, substitution was not found useful for comprehension.

In addition to the abilities to correctly decode sounds, recognize key words/phrases, and segment information, the successful listeners were found to use real-time assessment of input with a greater degree of success than the lower-scorers. In other words, the successful listeners were able to foreground and background sets of information correctly, according to their importance for the overall text meaning. As Goh

(2012) explained, listening is an online and cognitively demanding process, listeners may not be able to pay attention to every word or hold all pieces of information in mind while listening. They, therefore, need to be selective and purposefully opt to retain key information and ignore any which does not convey the text's main point. In this study, the high-scoring participants assessed pieces of information in real-time, based on their clear understanding of the pieces in their context (see Quotes 6.65-6.66). The low-scoring participants, on the other hand, assessed information based on their unclear understanding. They chose to foreground information without knowing exactly what it meant. They did so because they felt that it was necessary information for constructing the overall meaning of the text (see Quotes 6.69-6.70). It was unclear whether in the event real-time assessment of input was useful to the low-scoring participants because they also needed other strategies such as inferencing and elaboration to clarify their understanding. However, real-time assessment of input was shown to benefit the high-scoring participants since it helped decrease unnecessary cognitive demands by letting unnecessary information decay.

A strategy that was found to have impact on the use of real-time assessment, and which could cause misinterpretation, was prediction; in particular, when the prediction was not accurate according to the text's linguistic information. Prediction, in fact, was found to have both benefits and disadvantages, depending on the participants' background knowledge and familiarity with the text genre (discourse structure). In *Vitamin D* and *Talent*, which were both in genres familiar to the participants and had a main point that was explicitly stated at the beginning of the text, prediction was useful as it helped the participants to frame the texts' ideas very quickly and it directly steered their attention to the main point (Quotes 5.7-5.16). Prediction in *Hans Krebs*, which was in an unfamiliar genre for the participants, was not useful. Unless the participants verified their irrelevant prediction later while listening to the rest of the text, they were likely to misunderstand the main point. In *Hans Krebs*, for example, it was found that instead of understanding that people can become successful although they lack parental support, most participants understood that Hans Krebs became famous because of parental support (see Quotes 5.1-5.4). This misunderstanding can be explained by two reasons: 1) the listeners brought in incongruent background knowledge, namely that parents always encourage children to do great things in life, and 2) their linguistic processing (the lower-level processes) did not function effectively to detect words/phrases to realize that their prediction was wrong.

The lower-level processes were found crucial in understanding the texts which differed from participants' background knowledge. The analysis showed that most of the participants started off their listening by predicting what was coming up in the text. The participants then listened with some predicted information in mind (see Section 5.2.1). However, as prediction was made on initial linguistic input and background knowledge which was possibly not congruent with the text's linguistic information, the prediction could go wrong. If this was the case, the participants had to efficiently process the text at the lower-level in order to notice a shift or change in the text's discourse pattern. For example, in *Hans Krebs,* one participant (P3) realized on the basis of decoding the importance marker in the text 'a wonderful example' that the text was not organized in the way he predicted. Then he verified his prediction and because of this he was able to maintain his focus on the text's main point; otherwise, he would have missed it. Although some participants realized that their prediction was wrong, they could not decode the text quickly enough to identify the main point of the text and as a result did not understand the actual point intended by the speaker (see Quote 6.61). This finding provides support for Tsui and Fullilove's (1998) claim that when listeners' schema do not match the text schema, bottom-up processing (text processing based on linguistic information) is more important than top-down processing (text processing which relies partly on background knowledge) to fully comprehend the text. These findings also empirically illustrate Lynch's (2009) and Deroey's (2015) point that one element of linguistic knowledge that is vital for successful comprehension and has to be paid attention to in lecture listening, are importance markers, which speakers use to emphasize the importance of the points they make.

Global text comprehension, which means that the main point of the text is accurately and completely understood, is unlikely to be achieved without comprehension monitoring. It was found that the successful listeners continuously monitored their textual understanding throughout their listening process. They checked their on-going understanding of the listening passages. They evaluated a piece of information they did not comprehend according to whether it would contribute to their understanding of the entire passage and fixed it if it did. In addition, they verified a prediction when realizing it differed from the text. Comprehension monitoring also facilitated their use of real-time assessment of input. That is, it helped the participants decide whether they should foreground or background pieces of information, based on their text understanding. The findings thus empirically support O'Malley et al.'s (1989) and Goh's (2002) claim that comprehension monitoring is a useful metacognitive strategy in comprehending texts.

Comprehension monitoring and real-time assessment of input play an important role in text processing at the higher-level in overseeing text processing and in order to understand texts' 'true' meaning. In addition to these two metacognitive processes, the participants all employed two important cognitive strategies to bridge gaps in their text understanding, i.e. inferencing and elaboration.

With regard to inferencing, the results showed that every participant employed or attempted to employ it. This could be because the tasks required a summary of the input texts. Therefore, the participants might think that it was important to understand the global or discourse meaning of the texts and one way to achieve this aim was by making inferences on how the text was conceptually structured. Although, as acknowledged by the participants in the questionnaires, the ideas in *Talent* and *Vitamin D* were explicitly organized, some participants did not recognize the structure. They therefore made an inferential link between two independent ideas gained from the passages. As inferencing is a complex process that involves dealing with the various links being inferred between pieces of information, the participants with the ability to make inferences successfully could better see the structure of the text as a whole, resulting in deeper text comprehension. Inferencing is therefore one of the strategies needed for text comprehension.

It is worth pointing out that inferencing depended considerably on the results of text processing at the lower-level, i.e. linguistic processing, or on the words/phrases that the listeners obtained. Inferences were successfully made when they were based on key/necessary information and background knowledge which corresponded to the text's linguistic information (see Note 4.6 and Quotes 5.47). If this was not the case, the inferencing was not useful (see Quote 5.31).

In addition to inferencing, the participants were found to elaborate their understanding on the basis of their background or topical knowledge. Elaborations made on congruent background were found to be successful; otherwise, they were not. Successful elaboration benefited the participants as it created a platform to process the input text for overall understanding. In *Corruption*, the participants possessed background knowledge which was in line with the text (e.g., ‘corruption is a social problem’, ‘corruption has impact on public services’, and ‘corruption affects financial flow’). Their elaboration was therefore useful (see Quotes 5.47 and 5.50). In *Hans Krebs*, some participants were unsuccessful in their elaborations (see Quotes 5.52-5.53). This was because their background knowledge was not in line with the text. This finding contradicts the study by O'Malley et al. (1989) which showed that elaboration was useful by definition and contributed to listening success. The results from the present study, on the other hand, suggest that elaboration is useful when the participants have background knowledge which matches the text. However, when this is not the case and comprehension monitoring does not function properly, elaboration interferes with the listening process, resulting in misunderstanding texts.

The cognitive processes and strategies, as revealed in the data, were managed and controlled by metacognitive strategies, which were activated to make the cognitive processes and strategies function efficiently. In addition to the strategies discussed above, further strategies were employed by the participants, some of which appeared to be useful on their own, while some depended on others to be effective.

Pre-listening preparation was one metacognitive strategy that was used independently from the others and was beneficial to the participants. Before listening to the audio texts, the listeners engaged in pre-listening preparation and this helped them to clearly see the goal of their listening and in some cases reduce their listening anxiety. The use of this strategy is not likely to be task-specific, as once the listeners experienced a few tasks and were no longer nervous about their task performance, they were not evidenced using this strategy (see Section 6.2.1).

Directed attention was another metacognitive strategy that helped the participants to keep up with the focus of the texts after their attention had slipped away (partly because of unknown words or unfamiliar information). The benefit of this strategy, in itself, was not clear; it also depended on whether after the attention was redirected to the listening passage, the participants still managed to follow the text. If not, directed attention was unlikely to be useful. However, in terms of assessing language ability, the use of this strategy, to some extent, indicated limitations in lexical knowledge as it was activated mainly after the participants encountered unknown words and it was used more frequently by the low-scoring participants than the high-scoring participants.

Similar to real-time assessment of input, the effectiveness of selective attention was found to rely, to a large extent, on prediction. After predicting what they were going to hear, the participants were likely to pay selective attention to what they were expecting to hear. When the prediction was incorrect for the listening text and the prediction was not verified, selective attention was less likely to be useful.

In fact, the effectiveness of not only the metacognitive strategies but also nearly all the cognitive strategies, including inferencing and elaboration as discussed earlier, relies on other strategies in a very complex way. The inferences that the participants thought were correct according to the text turned out to be wrong when the inferences were not made on the key text message. The words recognized as key words were not key words because incorrect key information had been anticipated. For these processes to work effectively, other metacognitive strategies had to function effectively. In order to know whether the processes and strategies truly benefit performance, it is necessary to analyze them against task output. The fact that the strategies reported being used may not be used successfully on their own but may depend on others to be effective may explain why in studies such as Barkaoui (2014) and Purpura (1999) the frequency counts of the individual strategies used were not significantly related to task performance. These studies may have underestimated the interactive nature of processes and strategies.

To sum up, the findings discussed thus far suggest that text comprehension in listening-to-summarize tasks is an interactive process, relying on both cognitive and strategic processing abilities. At each level of cognitive processing, a number of cognitive and metacognitive strategies are employed to assist and manage the cognitive processes. In texts where the content is not congruent with listeners' background knowledge, listeners have to engage in comprehension monitoring and text processing at the lower-level (linguistic processing) effectively. Without this, they could easily misinterpret the text's main point. The test-takers used various processes and strategies which relied on each other in a very complex way to function effectively. Comprehension processing took place at higher-level processes, but the higher-level processes relied heavily on the effectiveness of processes at the lower-level and the use of cognitive (e.g., inferencing) and metacognitive strategies (e.g., comprehension processing, and real-time assessment of input). Without the application of these processes and strategies, the participants easily misunderstood the main point. Based on these results, this study thus recommends that a description of the listening construct of listening-to-summarize tasks should reference both the processes and strategies employed in listening.

9.2.2 Role of strategies in listening task performance

In addition to the question of what levels of cognitive processing the participants engaged in to comprehend listening input, the study also set out to investigate what strategies were employed to complete the tasks (RQ 1). The results, as provided in Chapters 5-6 and discussed earlier in this chapter, have shown that a number of strategies, including cognitive and metacognitive strategies, used by the listeners play an important role in success in listening comprehension. In addition, in line with previous research (e.g., Goh 2002), the results showed that while metacognitive strategies were necessary for task achievement, not all the cognitive strategies applied by the participants were useful.

Fixation and reconstruction are two cognitive strategies employed mainly by low-scoring participants and neither of them seemed to be helpful. Fixation, which was used mainly at the lower-level of text processing, did not appear to assist the low-scoring participants to grasp words any more correctly. Instead it diverted their attention from the listening text (see section 5.2.2). Likewise, reconstruction was unlikely to be helpful if the information reconstructed by the participant was incorrect (see section 5.2.5). Goh (2002) explained that such issues may occur because listeners struggle at the linguistic processing level (e.g., word recognition and parsing) with few chances to process the information at the higher-level and understand the global meaning of the text. This study also indicates that the use of fixation and reconstruction is associated with low(er) language knowledge.

The successful use of other cognitive strategies, i.e. inferencing, elaboration, and prediction, relied to some extent on other factors. Success in inferencing, in particular, depended on the effectiveness of text processing at the lower-level (linguistic processing). The participants who were successful in making inferences were able to recognize ‘key words’, which conveyed the key information of the text; without this, they were not successful (see section 5.2.3). This result empirically supports Field’s (2012) suggestion that inferencing is likely to be more successful when processing at the perceptual (decoding) and parsing stages functions effectively.

Success in prediction and elaboration, on the other hand, was found to depend mainly on the user’s background knowledge (whether it was in line with the text) and their familiarity with the text’s genre or discourse structure. When participants listened to a text in a familiar genre and on which they had relevant background knowledge, their prediction was likely to be accurate (see Quotes 5.5-5.14) and elaboration to be useful (see Quotes 5.46-5.51).

A comparison of cognitive and metacognitive strategies used by participants across performance levels revealed differences in strategy use. The high-scoring participants used more metacognitive strategies than the low-scoring participants. The low-scoring participants, on the other hand, used more types of cognitive strategies, some of which were not useful. The metacognitive strategies used by participants of all performance levels were pre-listening preparation and selective attention. Two metacognitive strategies, i.e., comprehension processing and real-time assessment of input, were employed more often by the high-scoring participants, whereas the low-scoring participants relied more on directed attention. In terms of cognitive strategies, the results showed that the low-scoring participants tended to rely on fixation and reconstruction. Two cognitive strategies used by participants from all performance levels are inferencing and elaboration. However, as discussed earlier, the high-scoring participants used these two strategies with a higher frequency and accuracy.

It has been suggested by Graham, Santos, and Vanderplank (2008) that more advanced listeners possess a larger linguistic base, and as a result have a larger information processing capacity, assisting them to have better control over their language use. In addition, advanced listeners are able to maintain attention or redirect it when distracted. Less skilled listeners are easily distracted when encountering anything unknown or unfamiliar and therefore need more strategies to fill gaps in their comprehension (Graham et al., 2008). If this is the case, the heavy reliance on cognitive strategies of the low-scoring participants, as found in this study, may be because they encountered more comprehension problems than the high-scoring participants. On the other hand, the high-scoring participants' use of fewer types of cognitive strategies but of more types of metacognitive strategies suggests that they processed the texts with less difficulty and as a result had better control over their listening process. The findings presented in the previous paragraph thus suggest that the pattern of strategy use can, to some extent, reflect the second language abilities of test-takers. Furthermore, with reference to the construct of listening-to-summarize tasks, as differences in strategy use are associated with language abilities and task achievement, this study, therefore, recommends that strategic processing ability should be recognised as one important component of the abilities assessed by listening-to-summarize tasks.

9.2.3 Effects of task modality (speaking and writing) on listening processing

Work in second language acquisition (SLA) (e.g., Bygate 1987) has revealed that modalities required after listening can differently affect task processing behaviours. In language testing in particular, Weir (2005) indicated that response types required in a test could possibly affect the way individuals perform the test. Consequently, this study also investigated the extent to which the different modalities (speaking and writing) required after the listening affected listening processing (RQ 1a). The participants in both the listening-to-speak and the listening-to-write tasks reported employing similar types of processes and strategies, but with different frequency. The similarity of the process and strategy types used in both tasks might be because the tasks required similar task output, i.e. a summary of listening. This result contradicts Bygate (1987) who has argued that the different nature of different modalities (e.g., speaking and writing) imposes different cognitive demands on language users, resulting in different cognitive processes while performing the language tasks. This could potentially be because in test tasks the purpose is much more explicitly and narrowly defined for the test-taker and the modality (speaking and writing) of task response therefore plays little role as far as listening comprehension processing is concerned.

However, one difference found between the tasks with different modalities concerned participants' perceptions of communication stress. More participants indicated feeling under more time pressure while listening-to-speak than while listening-to-write in all tasks. This result seems to support the idea that the tasks represent characteristics of real-world tasks, since real-world speaking typically takes place under time pressure and speakers formulate utterances instantly, while having little or no time to think about the content or check for grammatical accuracy and appropriacy (Bygate, 1987).

9.3 Role of listening comprehension and the listening abilities assessed by listening-to-summarize tasks

This section brings together the study's findings on test-takers' listening processing behaviours to discuss the role of listening and listening abilities assessed by the listening-to-summarize tasks and to address the overarching question: What listening abilities are assessed by EAP listening-to-summarize tasks? To begin with, the role of listening comprehension is discussed and then the listening abilities the tasks assess are presented.

9.3.1 Role of listening comprehension

The role of text comprehension in integrated tasks has been recognized to different extents. In some studies, scores obtained from integrated-test tasks were used to report test-takers productive skills (writing and speaking), without reference to comprehension abilities (see Frost, Elder, & Wigglesworth 2011; Gebril, 2010). In other studies (e.g., Gebril & Plakans, 2013; Sawaki et al., 2013), it has been indicated that integrated-test tasks do not only assess productive skills but also comprehension ability (reading and listening) which is reflected in task performance. Comprehension ability, as recommended by the latter group, should be acknowledged to be an integral part of abilities assessed by integrated-test tasks.

In this study, as discussed earlier in this chapter, listening comprehension played an essential role in the performance of integrated-tasks, i.e., listening-to-summarize tasks. The participants with different performance scores engaged in different listening processes and strategies with different degrees of success. The literature in language comprehension (see Anderson, 1985; Call, 1985; Færch & Kasper, 1986; Field, 2013; Rost, 2011; Vandergrift & Goh, 2012) has, in fact, acknowledged the complexity and interactiveness of the processes and strategies used in comprehension processing. However, to the best of this researcher's knowledge, comprehension processes and strategies, as presented in the literature, are described in a separate manner and no interactive models which display the interactiveness between processes and strategies and between cognitive and metacognitive strategies are provided. As far as listening comprehension is concerned, Field (2013), as presented in section 2.4.4, has described that listening cognitive processes are highly interactive; however, he does not explicitly illustrate how processes and strategies interacted during comprehension processing. Rost (2011) and Vandergrift and Goh (2012) have emphasized the roles of strategies in listening performance; however, they do not clearly show the interactiveness in the use of these strategies (see 2.4.2 and 2.4.3). To fill this gap, this section, based on the findings, presents a listening comprehension model which displays the interactiveness between the cognitive processes and strategies and between the cognitive and metacognitive strategies test-takers activated to understand listening input in listening-to-summarize tasks. Although described on the basis of the integrated-listening task findings, regarding the fact that the listening aim in this study is to comprehend aural texts' main point, this model should, to some extent, have implications for general listening with a similar listening purpose (understanding texts' main point).

Figure 9.1 illustrates the processes and strategies activated by the test-takers and the complex interactions between these processes and strategies. The green boxes at the centre of the figure show six cognitive processes activated by participants, the bottom three are lower-level processes and the top three are higher-level processes. The surrounding blue boxes are the cognitive strategies and the pink boxes show the metacognitive strategies used.

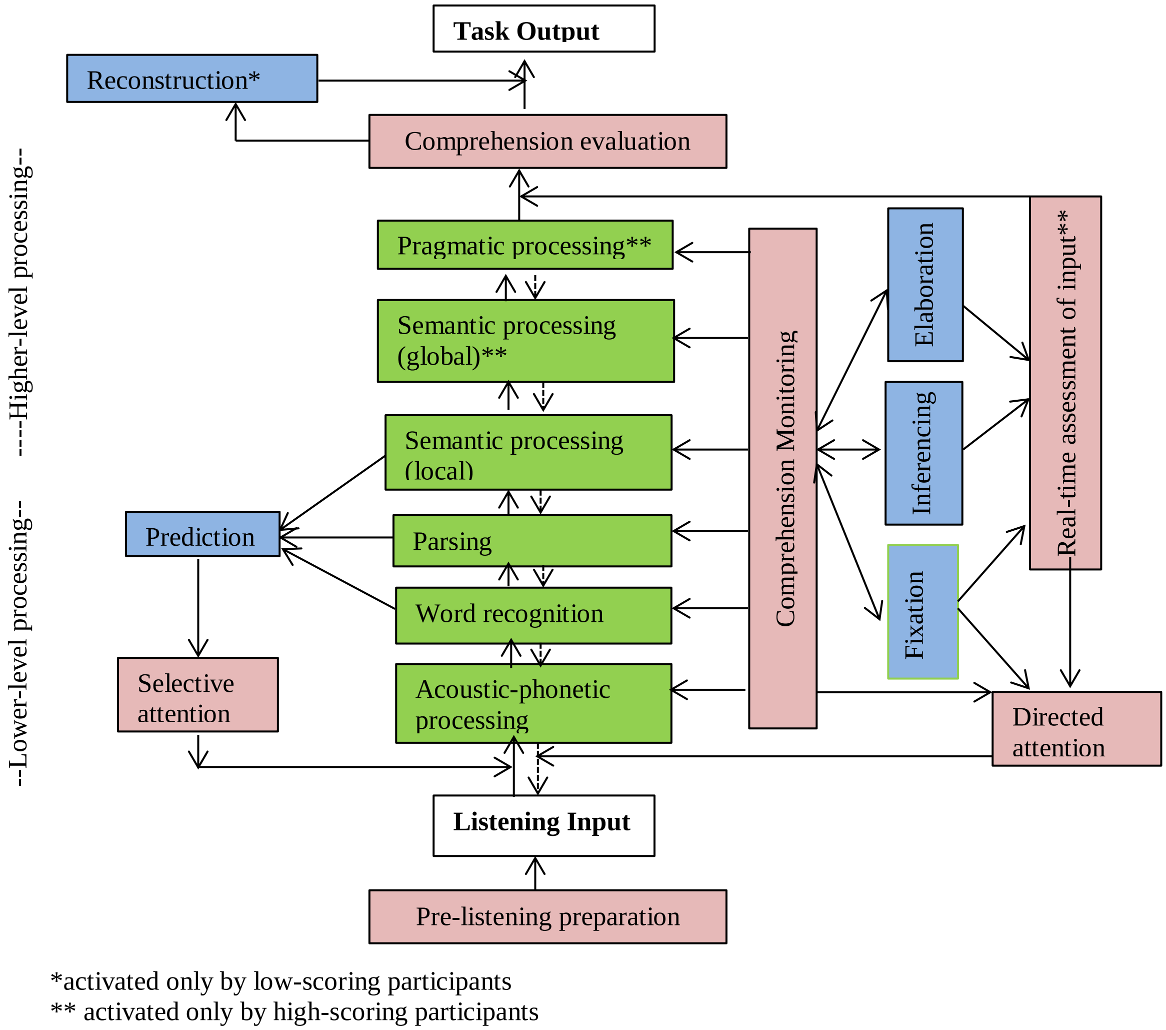


Figure 9.1: Listening comprehension model emerging from listening-to-summarize tasks

All the processes and strategies indicated in the Figure were used by the participants. Real-time assessment of input, semantic processing at the global level, and pragmatic processing were, however, adopted only by high-scoring participants. Reconstruction was used only by low-scoring participants to reconstruct listening ideas in order to fulfill task requirements. Although most of the processes and strategies were used by both high- and low-scoring participants, they were used with different degrees of success. The high-scoring participants activated three cognitive processes at the lower-level (acoustic-phonetic processing, word recognition, and parsing) with more automaticity and with higher degrees of success. Not only were they faster in word recognition and parsing, but also the words they recognized and chunks of information they parsed were more accurate than those obtained by the low-scoring participants and this enabled them to make more accurate inferences. The cognitive processes both at higher- and lower-processing levels were monitored by comprehension processing (a metacognitive strategy). In particular, this strategy was used to decide what strategies (fixation, inferencing, elaboration, and directed attention) should be activated to facilitate the six cognitive processes. The high-scoring listeners, however, used comprehension monitoring and real-time assessment of input more effectively that the low-scoring ones. The use of these metacognitive strategies was found very useful especially when the listeners' background knowledge was not congruent with the texts, particularly in order to verify that their prediction was wrong according to the text and to maintain their focus on the text's main point (see Chapter 6). Prediction and selective attention were commonly used by both low and high scorers, especially at the beginning of each listening to predict what they were going to hear next and to pay attention to such information accordingly.

The investigation of test-takers' processes and strategies in this study indicates that the listening-to-summarize tasks require listening abilities and that the different abilities that listeners have result in different levels of text comprehension and task performance. On the basis of these findings, this study, in accordance with Gebril and Plakans (2013) and Sawaki et al. (2013), shows that (listening) comprehension plays a genuine role in listening-to-summarize tasks. As a consequence listening abilities should be part of the description of the abilities assessed by integrated-skills tasks, i.e. listening-to-summary tasks, in addition to the productive skills involved.

9.3.2 Listening abilities assessed by listening-to-summarize tasks

To firm up the claim that listening comprehension plays a crucial role in integrated-skills tasks and that it should be recognized as part of the construct assessed by these tasks, this section presents the overall listening abilities required in the comprehension of academic listening input as operationalized in EAP listening-to-summarize tasks. Since in this study listening abilities were conceptualized in terms of cognitive and strategic processing abilities, the abilities are displayed in Figure 9.2. The Figure presents the three important components of integrated-test tasks – task input, comprehension processing, and task output (see 2.3) – and the abilities performed to comprehend task input (specifically in listening-to-summarize tasks) are presented in three categories – cognitive processes, cognitive strategies, and metacognitive strategies.

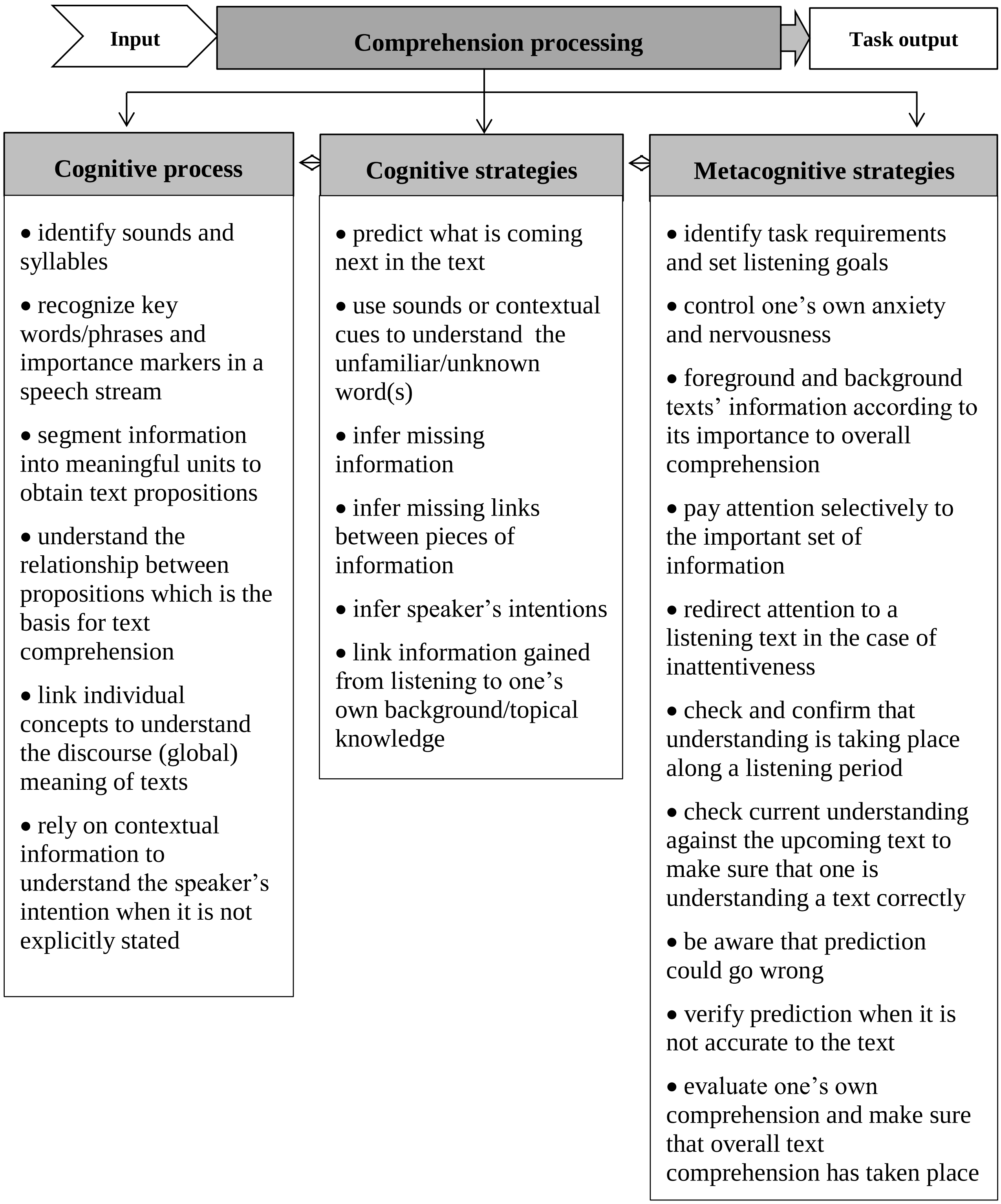


Figure 9.2: Listening abilities assessed by the listening-to-summarize tasks

Unlike Field's (2013) model of cognitive processing for listening, which is widely used in listening assessment research, the description of the listening abilities in this study includes the category of strategic processing as part of the listening abilities, i.e. cognitive and metacognitive strategies. Although Field's framework integrates some strategies such as monitoring and inferencing, it does not include some of the strategies which were found to benefit text comprehension in this study. In this study, successful listening, which means that the listener can correctly indicate the main point of the text in their summary, involved not only inferencing and comprehension monitoring but also other strategies such as real-time assessment of input, prediction, directed attention, and some compensation strategies such as approximation. This difference may partly be due to the source of data the researchers used. Field (2013) developed his framework based on the processes of L1 and expert L2 listeners' text comprehension processing. In this study, the data were obtained from the processing of L2 listeners who had more limited knowledge of the language as compared to those in Field (2013). This study, however, provides empirical support for Rost (2011) and Bachman and Palmer (2010), who stated that strategic processing should be included in the construct of language abilities.

9.4 Insights into the validity of listening-to-summarize tasks

As presented in Chapter 2, Weir (2005) stresses that test validation should not start from test developers' claim(s) but clear theoretical constructs that the tests are supposed to measure. In his socio-cognitive framework for test validation (which was employed in this study), Weir (2005) emphasizes that at an initial stage of test design and development, a test's theoretical construct has to be clearly defined and theory-based validity and context validity have to be justified before the test event. According to Weir

(2005), theory-based validity indicates that the abilities tests are theoretically expected to measure are relevant to those required in the target language use situation. Context validity, on the other hand, concerns the extent to which test tasks represent characteristics of real-world tasks. This section therefore discusses the results with respect to the theory-based validity of listening-to-summarize tasks as far as listening is concerned (or the extent to which the tasks efficiently tap into the listening abilities needed in the target situation), and the tasks' context validity (or the extent to which listening-to-summarize tasks represent characteristics of real-world tasks). This is to provide insights into the validity of these tasks.

The listening construct or abilities assessed by the tasks

Successful (academic) L2 listening, as presented in Chapter 2, involves a number of cognitive processes (both lower- and higher- levels) and strategies (both cognitive and metacognitive). Consequently, tests which are used to assess L2 listening abilities should tap into those processes and strategies in order to provide results that can accurately be generalized to real-life listening performance.

In terms of cognitive processes, a question is raised of the extent to which test tasks used are capable of tapping into high-level processes, especially semantic processing at the discourse level and pragmatic processing which are important for complete text understanding (see Field, 2012; Taylor & Geranpayeh, 2011). In this study, it was found that the listening-to-summarize tasks, which required the participants to produce a summary after listening, were able to do so. To be able to summarize the listening passage correctly, the participants had to understand not only the details of the passages but the overall meaning which was not always explicitly indicated and thus processing at higher-level was required.

Dividing task processing into two different levels, namely lower- and higher-level processes, this study found that the participants who successfully completed the listening-to-summarize tasks, meaning that they provided a correct main point and relevant supporting details, had engaged in higher-level processing. Their cognitive processes at the higher-level were, however, found to vary across the listening passages. In *Hans Krebs*, for example, the successful participant engaged in both semantic and pragmatic processing. This listening passage, as perceived by a large number of participants (see Section 8.6.2: cognitive processing demands), was cognitively complex. The majority said that they were unfamiliar with the content or could not predict the upcoming information. In addition, they agreed that more than one idea had to be attended to at a time and about one-third did not find the text explicitly organized or the main point clearly stated. To successfully comprehend this text, the participants had to infer links between ideas, connect ideas to the theme of the text, and rely on both the linguistic and contextual information to understand the point. In *Vitamin D*, on the other hand, the results showed that the participants successfully extracted the main point of the text correctly by relying on semantic processing at the global level, but not on pragmatic processing. This text, as compared to *Hans Krebs*, is less cognitively complex. About two-thirds of the participants indicated that they were familiar with the text content. In addition, the text processing was less likely to be cognitively demanding because they agreed that the ideas in the text were clearly organized and just one idea needed to be processed at a time.

Although some participants did not successfully engage in higher-level processes and did not obtain a full score on the summary content, they appeared to engage in parsing and semantic processing at a local level and obtained a score of '1' for content. Specifically, they were able to segment some key information correctly, understand some individual ideas, and integrate these ideas in their summary. These participants, however, were not able to include the main point of the text accurately. This was because, as revealed in the stimulated recall data, their inferencing and elaboration strategies were not successfully employed, resulting in the ineffectiveness of the use of comprehension monitoring.

The participants who scored '0' on the content were those engaging in lower-lever processes (acoustic-phonetic processing, word recognizing, and parsing), but not in higher-level processes as did the higher scorers, and their linguistic decoding was not as effective as that of the higher scorers. They were able to recognize words and parse chunks of information in the speech stream, but their perceived 'key words or key information' were incorrect and some of the key information was missing. To complete the task, they combined the words/phrases they had extracted from the passage with their background knowledge to reconstruct the key ideas of the text. As the key words and ideas they had extracted did not convey the key meaning, their reconstruction of the text was unsuccessful, resulting in an inaccurate summary and scoring '0' on content.

In addition to cognitive processes, successful understanding of input texts in listening-to-summarize tasks, as discussed earlier (see 9.2.2), requires the interactive use of a number of strategies. This includes the use of prediction, selective attention, directed attention, comprehension monitoring, inferencing, elaboration, real-time assessment of input, and comprehension evaluation. These strategies have been found by needs analysis studies of listening abilities in academic context (e.g., Imhof, 1998; Song & Cheng, 2006; Flowerdew & Miller, 2005) to be crucial for L2 listeners whose text processing is not as automated as L1 listeners. As discussed by Imhof (1998), the use of these strategies makes listeners become good listeners in an instructional setting. Flowerdew and Miller (2005) more particularly suggest that for successful academic listening, students should at least engage in high-level cognitive processes (semantic and pragmatic processing), directed attention, lowering anxiety, inferencing, performance evaluation, elaboration (on personal, academic, and world knowledge) and comprehension monitoring.

Success in real-life L2 listening, as Rost (2014) presents, is related to the engagement of processes and strategies in three processing domains – the cognitive, affective, and interpersonal domains. The cognitive domain encompasses all the complex skills of spoken language processing (linguistic, semantic, and pragmatic processing) that allow listeners to process quickly to keep up with the natural speed of proficient speech. The activation of cognitive processes found in this study constitutes the processing abilities in this cognitive domain. The test-takers' use of pre-listening preparation to get rid of text anxiety and nervousness, in addition, represents the use of strategies in what Rost (2014) calls the affective domain, which deals with stressful situations. In addition, in real-life communication, Rost (2014) describes that it is important that listeners are task-oriented and focus on the goal of tasks. They should have conscious strategies to monitor their own listening process to achieve listening goals. In this study, it was found that to complete the task requirement, which was summarizing listening input by speaking and writing, test-takers relied on several conscious strategies which are crucial for successful listening, including comprehension strategies, real-time assessment of input, and comprehension evaluation. These strategies, according to Rost (2014), are strategies in the interpersonal domain which enable listeners to achieve their goal.

Sawaki et al. (2013) suspect that integrated-skills assessment (listening-reading-writing) is capable of evaluating abilities that can be generalized to language use. If the generalized-language abilities are seen as abilities which are not task-specific and are likely to be transferred to other language use situations, then this study has shown that integrated-test tasks are capable of tapping into relevant abilities, such as the abilities to prepare for text processing (set listening goals, minimize listening anxiety, and get ready for listening) and to monitor the listening process (make sure throughout the task that comprehension takes place and solve comprehension problems if it does not). These abilities, as Cumming (2014) emphasizes, are basic abilities required not only to complete comprehension tests but also in academic contexts and the workplace.

Based on the results discussed, listening-to-summarize tasks have shown the potential to represent the construct of lecture listening in real-life situations. In particular, they have the potential to tap into the higher-level cognitive processes or semantic processing at the global or discourse level and pragmatic processing and strategies that are important to understand lectures outside test situations. However, it is important to note that the elicitation of some of the processes (such as pragmatic processing) and some strategies (such as real-time assessment) may be determined by the specific characteristics of input texts and task output.

Task authenticity and fairness

The second focus of the investigation was on task authenticity, fairness, and listening task difficulty, as perceived by test-takers. Authenticity and fairness were particularly looked into to investigate whether the test-takers who experienced the tasks will be directly affected by the use of the tasks. The results showed that the majority of participants strongly agreed that the tasks were authentic and a fair way of assessing their ability to use English for academic purposes. According to the participants, the tasks simulated what they encounter in their academic studies. This finding is consistent with the rationale underlying the use of integrated-test tasks, described by Cumming (2014) and Plakans (2015), namely, that integrated-skills tasks are capable of representing language tasks in real-life situations. Similarly, when asked how accurately the tasks were able to assess the individual skills (listening, speaking, and writing) involved in the task performance, over half of the participants (strongly) agreed that the tasks accurately assessed their listening, speaking, and writing abilities (see Table 8.4). This study thus supports the claims that integrated-test tasks, i.e., listening-to-summarize tasks, represent characteristics of real-world tasks and provide a fair way to assess language abilities. This helps support the claims that test results from these tasks are generalizable to the target language use domain.

Task difficulty and its relation to listening task performance

The study also investigated perceptions of listening task difficulty and their relation to task performance to see whether there was a systematic association between perceptions of listening task difficulty and listening performance. The results showed that text characteristics that made the listening input difficult, according to the participants, were information density (code complexity), unfamiliar listening content (cognitive complexity), and time pressure (communication stress). Of the four listening passages, *Hans Krebs* was perceived to be the most difficult, especially in terms of cognitive complexity (Table 8.10). The biggest number of the participants indicated that they were not familiar with the content or unable to predict what would come next. Because they were not familiar with the content and the text's information was dense, they had to pay attention to more than one idea at a time. These characteristics, as indicated by Lynch (2011) and Buck and Tatsuoka (1998), are in fact features of academic texts that listeners encounter in academic listening texts and which they have to be able to overcome if they are to process aural texts and participate in academic contexts successfully.

The perceptions were not very different when compared between the listening-to-speak and the listening-to-write tasks. The only exception was the perceptions of communication stress where a large number of participants indicated that they felt under time pressure while performing the listening-to-speak tasks. For the listening-to-write tasks, however, they remained neutral. This view, as discussed earlier in Section 9.2.3, may to some extent reflect the nature of real-world oral communication tasks; speaking tasks impose more stress on language users than writing tasks (Bygate, 1987).

The results on perceptions of task difficulty and their relation to the task performance indicate that, although the perceptions were found to relate to some aspects of task performance, no common patterns were found between the four listening texts investigated. The fact the *Hans Krebs* text was perceived as the most difficult was not, in any case, significantly related to task performance. The *Corruption* and *Talent* texts, with more or less the same difficulty as perceived by the participants, were either significantly or not significantly related to the performance in different respects. This may be due to two reasons. Firstly, it may be because the perceptions of task difficulty tend to relate to the listeners' topical knowledge rather than their experience performing the tasks. As observed in the stimulated recall, after listening to the first few sentences the participants judged whether they had topical knowledge and when they thought they did not have such knowledge, they stated that the task was difficult. Secondly, although the text was considered difficult, it was still possible for the listeners to successfully perform the task (providing an accurate summary) if their text processing at the lower-level, their metacognitive strategies, and especially their comprehension processing and real-time assessment of input functioned effectively. So although tasks were judged as being difficult, the listeners might still be able to do well if their cognitive processing worked effectively. In other words, perceptions of difficulty may not determine task achievement, but task processing abilities are likely to do so. These findings contradict SLA studies such as Robinson (2001) and Tavakoli (2009), which suggest that perceptions of task difficulty related to task performance. These findings, however, support Elder et al. (2002) who stated that test-takers' perceptions of task difficulty may not be such a useful source for describing task difficulty for the listening construct underlying integrated-listening task. As explained by Elder et al. (2002), the different findings that emerge in teaching and testing contexts may be because under testing situations, test-takers may concentrate very hard on providing accurate performance/responses and may not pay attention to task characteristics (e.g., familiar/unfamiliar lexis and less complex/more complex structures) and conditions (more cognitive/less cognitive demands) offered to make tasks easier or more difficult.

In sum, this investigation of test-taker's listening processing supports the claim that listening ability should be recognized to be part of the construct underlying listening-to-summarize tasks. This is because the tasks were found to require listening abilities to comprehend input materials and this understanding is a prerequisite for task production in integrated listening-to-summarize tasks (written or spoken summaries). This investigation, together with that into the test-takers' perceptions of task authenticity, fairness, and listening task difficulty, furthermore, support the use of listening-to-summarize tasks in assessing English for academic purposes. The tasks have the potential to capture the abilities required in real-life situations, especially listening cognitive processing at a high level and the use of metacognitive strategies. The characteristics of the tasks and listening task difficulty as perceived by the test-takers, to a large extent, represent some of the tasks they encounter in their academic studies, and these tasks – as the test-takers perceived – accurately reflect the abilities they need to perform in their academic studies. From this perspective, the use of listening-to-summarize tasks that reference to both receptive (listening) and productive (speaking/writing) abilities in their construct) to assess academic English abilities should therefore be justified.

9.5 Summary

Based on the results and the discussions provided, this study suggests that listening comprehension abilities should be a recognized part of the construct underlying integrated-listening tasks in addition to the productive skills involved. The study shows that test-takers engaged in a number of processes and strategies to comprehend listening input text. In addition, it reveals differences in the processing (in both processes and strategies) used when comparing between different performance levels. As a result, it is recommended that these processes and strategies are part-and-parcel of the description of the listening construct of listening-to-summarize tasks. With regard to test-taker perceptions, the findings on task authenticity and fairness are consistent with previous research. Namely, the tasks, as perceived by the test-takers, represent real-world tasks and accurately tap into the language skills that need to be performed in an academic context. However, the investigation of the relationship between perceptions of listening task difficulty and task performance did not point to any common patterns. This seems to suggest that perceptions of task difficulty may not be a useful source for test validation. This may be because 1) participants' judgments of listening task difficulty were mainly based on topical knowledge rather than their actual experience of the listening task performance, and 2) the determinant of successful listening is likely to be text processing at the lower-level and strategy use may affect task performance more than test-takers' perceptions. Based on the findings, this study supports the usefulness of listening-to-summarize tasks in assessing English for academic purposes.

The implications of these findings will be offered in the next chapter. In addition, limitations of the study will be pointed out and recommendations will be made for further research.

Chapter 10 Conclusions

10.1 Introduction

This chapter concludes the study by first providing a summary (10.2). This will include restating the research aims and the research questions, and summarizing the methodology and the main findings. Next, the contributions and implications of the study are discussed in 10.3 and 10.4, respectively. The final section (10.5) acknowledges the limitations of the study and provides directions for future research.

10.2 Summary of the study

A major concern in language assessment is the extent to which tests tap into the abilities to use language beyond the test situation. Since acts of real-life communication often involve several language skills, integrated tasks are increasingly used in language tests. These tasks require test-takers to use at least two language skills (e.g., listening-speaking; reading-writing) and are perceived to better represent the characteristics of real-world tasks (Cumming et al., 2004; Plakans, 2009; 2014). Studies investigating the abilities actually assessed by integrated tasks, however, have primarily focused on the productive skill(s) involved (speaking and writing); the role of receptive skills in integrated task performance is still unclear, particularly for tasks involving listening. In fact, the validity of language tests depends, to a great extent, upon a clear test construct. When the construct is not clearly defined, it is difficult for testers to support interpretations and decisions made on the basis of test scores. With this in mind, this study set out to investigate the listening construct underlying listening-to-summarize tasks.

Drawing upon the literature in language testing (see Chapter 2), this study conceptualized the construct or abilities assessed by language test tasks as the cognitive and strategic processing test-takers engage in during task performance and explored the listening construct of the listening-to-summarize tasks by addressing the following research questions.

The overarching question was "What listening abilities are assessed by EAP listening-to-summarize tasks?" and the subordinate questions were formulated as follows.

1. What cognitive processes and strategies do ESL test-takers engage in while performing (adapted) PTE Academic listening-to-summarize tasks?
    a. Are there any differences in the processes and strategies when compared between tasks with different language modalities, namely speaking and writing, and between performance levels?
2. What are ESL test-takers' perceptions of (adapted) PTE Academic listening-to-summarize tasks and of listening task difficulty?
    a. Is the listening difficulty as perceived by the test-takers related to their listening performance?

The research data comprised 1) test-takers' cognitive and strategic behaviours in processing the listening input of (adapted) PTE Academic *Re-tell Lecture* and *Summarize Spoken Text* tasks, and 2) their perceptions of the tasks and of listening task difficulty and the relation of this difficulty to task performance. To answer the first set of questions, 12

Thai ESL learners pursuing post-graduate level degrees at a UK university completed four listening-to-summarize tasks with different listening inputs. Two tasks required an oral summary of the listening and the other two a written summary. After completing each task, the participants conducted stimulated recalls which required them to watch a video recording of their own behaviours and explain what they were doing or thinking about while listening. The participants' notes taken during task completion and their oral/written content summaries were analysed to supplement the stimulated-recall data. By using a coding scheme developed on the basis of existing listening frameworks (see Appendix 5), the data were analysed by two coders.

To answer the second set of research questions, on the perceptions of tasks and listening task difficulty and their relationship to task performance, another group of 60 Thai students studying at four universities in the UK was included. After they completed each of the four listening-to-summarize tasks on a one-to-one basis, a perception questionnaire was administered. The data were quantitatively analysed using the statistical software SPSS, running descriptive statistics and correlations.

In what follows, the findings are summarized according to the research questions.

Question 1: Overall processes and strategies used to comprehend the listening input in the listening-to-summarize tasks

The analyses showed that the participants relied on both cognitive processes and strategies to understand the listening input. The processes comprised six cognitive processes, categorised into two levels of processing: three lower-level processes (acoustic-phonetic processing, word recognition, and parsing) and three higher-level processes (semantic-processing at the local level and at the global level and pragmatic processing). The lower-level processes were engaged in by all participants while listening to almost every passage. Although some participants were not aware of their activation of acoustic-phonetic processing and word recognition, their engagement could be inferred from what they described knowing or understanding while listening in their stimulated recalls in combination with their notes and summary content. The number of the participants engaging in the higher-level processes was smaller than those who engaged in the lower-level processes. To produce a good summary and receive a full mark on content, the participants had to engage in semantic processing both at the local and the global levels and pragmatic processing, except for the tasks with *Vitamin D*, where the participants could obtain a full mark without pragmatic processing. The perception questionnaires revealed that this is because the text input of this particular task was clear and easy to understand. Therefore, the participants may not have needed to rely on contextual information to understand the entire meaning.

In addition to these processes, the participants used cognitive and metacognitive strategies in their listening tasks. Five cognitive strategies reported by the participants were: 1) prediction, 2) fixation, 3) inferencing, 4) elaboration, and 5) reconstruction. Fixation and reconstruction were the two cognitive strategies used mainly by the low-scoring participants. They were found to be unlikely to be useful to the whole listening process since these participants did not provide accurate information after fixation or reconstruction. Inferencing was used by every participant and hence was the most frequently used cognitive strategy. Inferencing was however not always successful as it depended, to a large extent, on text processing at the lower-level or on linguistic processing. When inferences were not made on key or necessary information of the text, the inferred meaning of the text tended to be wrong. For inferencing to function effectively, it was important that the lower-level processes (decoding/ parsing information necessary to understand the main point of the text) were effective. Success in prediction and elaboration depended on several other factors including the participants' topical knowledge, familiarity with the text's discourse, and metacognitive strategies. Both prediction and elaboration were based on the listeners' topical knowledge and familiarity with the text type. When such knowledge was not congruent with the texts' linguistic information, the use of these strategies was less effective. Unless the participants were able to recognize some importance markers in the text, realize their predicted and elaborated information could be wrong, and monitor their comprehension effectively, they misunderstood the whole meaning of the text.

Six metacognitive strategies were reported by the participants, namely 1) preparing for listening, 2) selective attention, 3) directed attention, 4) comprehension monitoring, 5) real-time assessment of input, and 6) comprehension evaluation. Three strategies used by a large number of participants were comprehension monitoring, selective attention, and directed attention. Success in using each of these strategies however depended on the activation of other strategies and processes. An exception is the preparing for listening strategy, which was useful in itself. It occurred before the audio texts started, when the participants were setting their listening goals, and reduced their listening anxiety. Selective attention was found very useful when the listeners' predicted information was similar to the textual information. However, since selective attention is about purposefully and primarily focussing on the predicted key point, it led the listeners to focus their attention on the wrong point when their prediction was not accurate, unless their linguistic processing worked effectively. In addition, it was found that the effectiveness of comprehension monitoring relied on lower-level processes, and real-time assessment of input and directed attention depended on comprehension monitoring to be successful.

The levels of cognitive processes for listening tapped into by the tasks

The results showed that listening-to-summarize tasks are capable of tapping into higher-level processes. Listeners who were able to provide an accurate summary, including both accurate main point and relevant details, engaged in semantic processing both at the local and the global levels for all four passages used as listening input. Pragmatic processing was also necessarily conducted by the participants to fully understand three of the input materials, namely *Corruption*, *Hans Krebs*, and *Talent*. Only in *Vitamin D*, was there no evidence that pragmatic processing was needed to provide a good summary. This might be because the text itself was not linguistically difficult for the listeners and the main point was explicitly indicated. In all four listening texts, the participants who were found to engage in semantic processing at the local level but not the global level were able to provide a number of relevant details, but not an accurate main point of the listening input despite trying to do so. This is partly because the participants were not able to notice and decode the importance markers in the text correctly and, although they understood separate ideas in the texts, they were not able to correctly link them to the theme of the text to understand its actual meaning. The listeners who operated only low-level processes were able to provide a few pieces of correct information but included some irrelevant or wrong information in their summary; as a result, they scored low on content. This is partly because their lower-level processes, especially word recognition and parsing, did not function effectively and they had to rely on fixation and directed attention to basically understand the literal meaning and complete the tasks.

Based on the results, it can be concluded that the listening-to-summarize tasks, which require test-takers to summarize listening input, tap into both lower- and higher-level listening processes. Three important characteristics of the listening input that appeared to force text processing at the higher-level, as found in this study, are 1) no explicitly stated main point, but one that has to be inferred from the key information, 2) the use of importance markers to signal the important piece(s) of information, and 3) unpredictable discourse structure for the listeners.

Question 1a: The differences between the strategies used across performance levels and task modalities

The results in relation to the use of strategies pointed to a slight difference when comparing between the tasks with different modalities; the high scorers relied more on comprehension processing, real-time assessment of input, and inferencing than the low scorers. The low scorers, on the other hand, depended more on fixation and reconstruction. The comparisons between the cognitive and metacognitive strategies used by the listeners with different performance levels showed differences in terms of frequency and types. While the high scorers used more types of metacognitive strategies, the low scorers used more types of cognitive strategies. The five cognitive strategies identified were all used by the low scorers in each listening passage, with a high frequency of fixation, elaboration, and reconstruction. The high scorers, in contrast, relied mainly on inferencing. With regard to the use of metacognitive strategies, it was found that the high scorers reported using all types of metacognitive strategies, with a high frequency of comprehension monitoring. The low scorers were found to use fewer types of metacognitive strategies, with a high frequency of directed-attention.

Different types of strategies were used with different purposes. The use of cognitive strategies was to solve listening problems and fill gaps in listening comprehension; a more extensive use of cognitive strategies could thus indicate more gaps in linguistic knowledge and hence in listening comprehension. Less use of cognitive strategies and more use of metacognitive strategies by higher-scorers may therefore indicate that they experienced less listening difficulty and hence they had better control over their listening process. Based on these results, this study concludes that in addition to cognitive processes, strategic processing behaviours play an important role in task performance and thus they should be referred to in the listening construct underlying listening-to-summarize tasks.

Question 2: Test-takers' perceptions of tasks and listening task difficulty

The results regarding the perceptions of tasks showed that the majority of participants (strongly) agreed that listening-to-summarize tasks represent the tasks they encounter in their academic life and fairly assess their language ability. A higher number of participants agreed that the tasks accurately assessed their listening ability than of those indicating for speaking and writing abilities.

In terms of listening task difficulty, investigated via the perceptions of code (linguistic) complexity, cognitive complexity, and communication stress, the results showed variation in listening task difficulty. *Hans Krebs* and *Corruption* are perceived to be more difficult than *Talent* and *Vitamin D*. Although a higher number of participants agreed that the textual information of all listening passages was dense, they did not agree that *Corruption* and *Hans Krebs* were clearly organized or that they were familiar with the ideas in these two texts. Hence, *Hans Krebs* and *Corruption* are likely to be more linguistically and cognitively complex than *Vitamin D* and *Talent*. In addition, although about half of the participants indicated that they were familiar with the task type (listening to an academic lecture and summarizing it), they reported being under time pressure, especially while performing the listening-to-speak tasks.

Question 2a: The relationship between perceptions of listening task difficulty and listening performance

The investigation of the relationship between listening task difficulty, as perceived by the participants, and listening task performance indicated that the perceptions were significantly related to task performance in a few respects. However, no common patterns of correlations were identified. The perceived difficulty of the *Hans Krebs* tasks was not, in any respects, significantly related to listening performance (assessed via summary content). The perceptions of the difficulty of *Corruption* were negatively and significantly correlated to the content score in the listening-to-speak task, meaning that the participants who perceived the listening text as linguistically complex and cognitively demanding provided a less accurate summary. Additionally, fluency was negatively and significantly related to cognitive complexity, indicating that the participants who thought the task was cognitively demanding were likely to speak less fluently. However, these correlational patterns were not found in the tasks with *Talent* although the listening task difficulty, as perceived by the participants, was more or less the same as that found in *Corruption*. In *Vitamin D*, which was perceived as the easiest listening passage among the four passages, only the perceptions of cognitive complexity were negatively related to the content summary. This shows that the participants who thought the text was cognitively demanding were likely to provide a less accurate summary of the text.

As discussed in Chapter 9, the fact that no common patterns of relationship between perceptions of task difficulty and task performance were identified could be due to two reasons. Firstly, perceptions of task difficulty tended to relate to the listeners' topical knowledge rather than their linguistic ability and actual experience in listening to the text. Secondly, although the text was considered difficult, it was possible for the listeners to successfully perform the task (providing an accurate summary) when their cognitive processes at lower-level and metacognitive strategies, especially comprehension processing and real-time assessment of input functioned effectively.

10.3 Contributions of the study

10.3.1 Theoretical contributions

This study provides important insights into the theoretical construct underlying integrated-listening tasks in a number of ways. Weir (2005) has indicated that having clear theoretical constructs of language test tasks is important for test designers to select tasks that best suit their testing purpose and later to justify the interpretations of test scores and use. However, as far as integrated-listening tasks are concerned, it has so far been unclear whether and to what extent listening played a role in task performance and what listening abilities were assessed by this task type. Drawing on the language testing literature, listening abilities in this study were conceptualized as the cognitive and strategic processing behaviours listeners activate to comprehend the listening input. Adding to previous research in this area, this study has revealed the set of abilities required in listening comprehension. Namely, what is crucial for successful comprehension are effective processing at the lower level (especially the ability to recognize importance markers), higher-level processes (i.e. semantic and pragmatic processing), and the use of such strategies as inferencing, comprehension processing, and real-time assessment of input. These processes and strategies, as discussed in Chapter 9, work interactively and depend on one another in a very complex way to bring about complete text understanding.

This study also extends our understanding of the association between task processing behaviours and task performance. The investigation of the processes and strategies the participants engaged in in this study showed that the processes and strategies function interactively and depend on one another in a very complicated way. Successful inferencing which facilitated semantic and pragmatic processing was found to rely on text processing at the lower level, such as word decoding and parsing. Real-time assessment of input, on the other hand, relied on comprehension monitoring. Selective attention was effectively used when prediction was accurate according to the text's linguistic information. When prediction was wrong, it was important that comprehension monitoring functioned effectively with the help of text processing at the low or linguistic processing-level. If this was not the case, texts could be misinterpreted. The fact that several strategies reported by the participants were not used successfully on their own, but relied on other strategies and processes to be successful, could possibly explain why the frequency count of individual strategies used in previous studies, such as Barkaoui (2014) and Purpura (1999), was not related to task performance. In such a case, as discussed in Chapter 9, it may that the interactive nature of processes and strategies may have been underestimated.

Based on the findings of test-takers' listening comprehension processing, this study has formulated a listening comprehension model, which helps fine tune existing cognitive frameworks in several ways. First, it indicates that strategies (both cognitive and metacognitive) play important roles in (L2) listening performance. Since some of these strategies (such as prediction, selective attention, and real-time assessment of input) have not been explicitly indicated in Field (2013) – the cognitive processing framework for listening being used in L2 listening assessment – this indication seems to be crucial. This model in addition illustrates the interactiveness between cognitive processes and strategies and between cognitive and metacognitive strategies test-takers activate to understand listening input in listening-to-summarize tasks. This interactiveness has in fact been acknowledged in several listening frameworks such as Field (2013), Rost (2011), and Vandergrift and Goh (2012); however, none of these models clearly displays it.

This study in addition contributes to knowledge in that it has shown that listening-to-summarize tasks, both listening-to-speak and listening-to-write tasks, are able to tap into higher-level processing, which is aimed to be assessed by several testing situations. In the tasks where test-takers were required to summarize a listening passage, the high- and the low-scoring listeners were found to rely on different levels of text processing to comprehend listening input. Although lower-level processes were found important to process text for comprehension, the sole activation of these processes proved insufficient to enable test-takers to provide an accurate summary of the listening input; such processes enabled the listeners to understand only the literal meaning of the texts. This was the case for low-scoring participants, who were able to include in their summary only chunks of relevant information, but not the texts' main point. What was crucial for test-takers to be able to provide a correct main point was successful engagement in high-level processes, i.e., semantic processing at the global (discourse) level and pragmatic processing. This was evidenced in the high-scorers' task-processing behaviours.

The final theoretical contribution that can be drawn from the findings is related to the role of test-takers' perceptions in test validation. Test-takers' perceptions were found to be useful for indicating task authenticity and fairness in this study. Considering test-takers as a useful data source as they have experienced performing the tasks, the study investigated task authenticity, fairness, and difficulty via test-takers' perceptions. The majority of participants strongly agreed that the test tasks represent those they have come across in their real life and are a fair way to assess their English abilities. These agreements can, to some extent, support the generalization of test scores beyond the test situation. The investigation of perceptions of listening difficulty, on the other hand, has pointed to the characteristics of the texts that increased listening difficulty such as information density and text discourse structure. These characteristics, as revealed in qualitative data, required test-takers to activate some particular processes and strategies, such as semantic processing at the discourse level and real-time assessment of input to listen to the texts successfully. The correlation analyses, however, did not indicate common patterns of relationship between the perceptions of task difficulty and task performance. The listening input which was considered difficult by the participants was successfully comprehended by the listeners if their text processing at the low-level functioned effectively. Based on the findings, this study therefore concludes that test-takers perceptions are useful to indicate levels of task authenticity, fairness, and the characteristics of tasks contributing to task difficulty. This source of data, however, may not be useful to estimate overall task difficulty since the participants who perceived tasks as difficult either succeeded or failed the tasks, depending on their processing abilities.

10.3.2 Methodological contributions

This study shows the importance of mixed methods in investigating test-takers' processing behaviours. The study set out to investigate test-takers' cognitive and strategic processing behaviours by using stimulated recalls, where the participants were required to explain what they were thinking about or paying attention to while listening. Although this data collection method proved to be useful as the recalls provided insights into the processes and strategies participants engaged in, some processes and strategies went unrecognized by the participants, especially automated processes. To bridge this gap, the participants' notes and their content summaries were also included as research data and analyzed to complement the stimulated recall data. The analyses of the notes and the summaries did not only point to the presence of automated listening processes that participants were unable to recognize or report but also indicated the extent to which the reported processes and strategies were conducted successfully. Some of the data from these three sources triangulate with each other in indicating the activation of some processes (semantic and pragmatic processing) and strategies (fixation, inferencing, and elaboration). Based on these findings, it is advisable to analyze test-takers' notes (if any) and the content of task output to complement and triangulate verbal data. This is in particular to study task processing behaviours.

10.4 Implications of this study

10.4.1 Implications for test developers

Some language testing researchers have warned against the use of integrated-skills tasks to assess language abilities because of their confounding effect, or the problem that the ability to perform one skill may have an impact on the performance on another skill required by the tasks. For example, listening ability may influence speaking or writing ability in integrated listening-speaking and listening-writing tasks. However, in this study, the participants judged that this task type represents the tasks that they encounter in their academic studies and fairly assesses the language skills involved in task performance. In addition, the tasks, as discussed earlier, assessed the listening comprehension abilities required in real-life situations. Since real-world communication generally involves a combination or integration of language skills and this task type has proved to assess the abilities performed in authentic language use, the above concerns do not seem justified and it is meaningful to use integrated-skills tasks in the assessment of language ability.

It was found in this study that listening comprehension involved a number of processes and strategies and the performance of these processes and strategies was found to differentiate intermediate listeners from advanced (see Chapter 7). Listeners with different performance levels differed in the processes and strategies used to comprehend the listening input. It is thus advisable that where integrated-skills tasks are used for assessing language ability, the comprehension ability should be explicitly scored via the content of productive tasks. Furthermore, given the crucial role of comprehension ability, it seems justifiable to make the weight of the content score up to half of the total score in integrated-test tasks and let criteria specifically related to the productive skill (e.g., fluency, linguistic accuracy and range) contribute the other half.

In terms of content scoring, despite the fact that complete text understanding and correct content summaries are the target, this study also recommends considering partial understanding and partly correct summaries as an indication of listening abilities. While partial understanding is clearly at the lower level and therefore not deserving of high scores, giving no points also does not accurately represent test-takers' abilities in task performance. Although less successful listeners did not engage in high-level processes and strategies as successfully as the successful listeners when providing the correct main point of texts, they have been revealed to possess some language knowledge and processing abilities which form a basis for text understanding. For example, test-takers who can provide most of the relevant ideas in their summary but not the main point have been shown to be capable of processing texts linguistically and semantically at the local level. Test-takers who provided accurate propositions/chunks of information engaged in linguistic processing. The scoring criterion should thus take into account the processes and sources of knowledge that test-takers use in providing the content and allocate scores accordingly.

An implication for test developers who aim to reveal the construct underlying test tasks is that it is necessary to look into test-takers' processing behaviours as they point to the different processes and strategies activated to complete tasks, their strengths and weaknesses, and the extent to which such processes and strategies are used successfully. In addition, it is important that both verbal data and analyses of task performance are included in such investigations since it was found that in some cases the reported strategy use did not link to successful use. Only relying on reported usage may give a wrong picture of strategy use and the abilities performed in task completion. Therefore, to compensate for these issues of the veridicality (the accuracy of verbal data) and comprehensiveness of verbal report methods, this study recommends looking into additional sources of data, such as learners' notes and the content of task output to obtain a complete picture of the abilities assessed by the tasks investigated.

The bases of tasks investigated in this study were originally *Re-tell Lecture* and *Summarize Spoken Text* items. Although these tasks were adapted to suit the research purpose, the task adaptation (see 3.4.2) did not change the nature and the main requirements in task performance (listening-speaking and listening-writing). It is thus possible to provide the implications for PTE Academic in terms of its theory-based and context validities. The findings of the task processing behaviours show that these (adapted) tasks tap into cognitive processes and strategies (cognitive and metacognitive strategies), which, as discussed in Chapter 9, are required for success in real-life academic listening. These results, to some extent, indicate the theory-based validity of the *Re-tell Lecture* and *Summarize Spoken Text* items. In addition, the findings of test-takers perceptions show that the majority of test-takers in this study perceived the tasks as authentic. The participants (strongly) agreed that the tasks represent the characteristics of tasks they encounter in academic contexts. In addition, the tasks, as perceived by the test-takers, assess their academic English abilities fairly. These findings, to some extent, show the context validity of the (adapted) *Re-tell Lecture* and *Summarize Spoken Text* items. Therefore, it can be concluded that the findings support the use of these items in the PTE Academic which aims to assess English for academic purposes.

10.4.2 Implications for language teachers

This study has implications for teachers who are responsible for teaching academic L2 listening. First, it provides them with conceptual information of what abilities are involved in academic L2 listening overall, including both cognitive and strategic processing abilities and the interaction between these processes and strategies. In addition, given that listening in real-life situations does not occur in isolation but in combination with other language skills, it is also important to suggest that listening should be practised in combination with other skills such as speaking and writing in order to fully simulate real-world situations of language use. Practising language skills in combination, as suggested in Chapter 2, is key to succeeding in language learning.

The study revealed aspects of language that learners, especially at the intermediate or upper-intermediate levels who have acquired a certain level of language knowledge, need to learn in order to be successful in academic listening. What were found to play an important role in understanding the main points of lectures are the discourse structure and importance markers used by the speakers. Although the listeners in this study were likely to be familiar with ‘traditional’ logical connectors and linking words and be able to recognize such markers as ‘first’, ‘then’, and ‘however’, (given their overall proficiency level) most of them struggled more with importance markers which take the form of phrases and which are expressed in a more passage-specific manner such as ‘it does that by accident’, and ‘Krebs is a wonderful example to me of how’. Inability to recognize the importance markers used and how the independent ideas in the text are organized may cause misunderstanding of lectures’ main points. Given that these types of markers are common in academic lectures (Lynch, 2011; Deroey, 2015), it is important that classroom teaching for academic listening emphasizes and helps learners develop recognition and comprehension of this marker type.

What is also important for listening to academic lectures are metacognitive strategies. Some listeners were found to rely on their background knowledge to comprehend the texts without realizing that it was incongruent with the text’s linguistic information. As a result they misunderstood the main point of what they were listening to unless their comprehension monitoring and real-time assessment of input worked efficiently. Thus, in addition to emphasizing the importance markers or discourse markers of any type, it is recommended to enhance learners’ metacognitive strategies, especially the ability to monitor their own comprehension (e.g., evaluating their on-going understanding and questioning the relevance of their background knowledge to listening content) and real-time assessing of the input.

10.5 Limitations and future research

Despite having been carefully designed, this study has some limitations. One is related to the research tasks. The study was limited to only listening-to-summarize tasks, requiring oral and written summaries after listening. Integrated listening-test tasks with different requirements for task output (e.g., retelling or discussing the listening texts) were not the focus. However, since these have not been looked into, it is unclear to what extent the present study's findings generalize to those tasks, particularly since some strategies used by the participants, such as reconstruction, selective attention, and real-time assessment of input, were found to be task/item-specific. In addition, the tasks in this study were completed by a group of intermediate and upper-intermediate Thai ESL learners in the UK. Their background and topical knowledge was found to play a crucial role in their cognitive processing and their use of strategies. Consequently, the findings may not be generalizable to other groups of L2 listeners with different background profiles. Hypothesizing that educational experience and strategy training may affect the way listeners developed and use processes and strategies, this study proposes a replication in different contexts and with participants with different educational and language backgrounds.

Another limitation of this study relates to the fact that task performance was primarily investigated with reference to listening comprehension only. This was motivated by the lack of attention in the literature to this receptive skill in integrated tasks, and in particular the unclear role of listening comprehension in integrated-skills tasks. While one group of researchers found that comprehension ability played a distinct role and recommended that this ability be explicitly referenced in the construct underlying integrated-test tasks, others have concluded that integrated-test tasks (e.g., reading-writing) are the same as independent skill tasks (e.g., writing) in terms of what they measure. In this study, it was found that listening comprehension abilities play an essential role. High and low scorers differed in their cognitive and strategic processing behaviours, suggesting that they possess different listening comprehension abilities. This finding suggests that comprehension ability exists and plays a role in successful task performance. However, what this study did not explore in detail was the productive performance on the task and the links between comprehension and production. This study thus recommends further research that investigates both the comprehension and production parts of the task processing and performance to gain additional insights into the interaction between these and to obtain a fully comprehensive picture of the abilities that are actually assessed by integrated-skills tasks as a whole.

List of references

Alderson, J. C. (1984). *Reading in a foreign language: A reading problem or a language problem*. In Alderson J., Urquhart A. (Eds.), Reading in a foreign language (pp. 1–27). London: Longman.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Anderson, J. R. (1985). *Cognitive psychology and its implications*. New York: Worth Publishers.

Asencion, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners*. Northern Arizona University. ProQuest Dissertation & Theses (PQDT).

Asención Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes, 7*(3), 140-150.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing, 19*(4), 453-476.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Asessment Quarterly, 2*(1), 1-34.

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

Bachman, L. F, & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: an empirical study of their veridicality and reactivity. *Language Testing, 28*(1), 51-75.

Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL-iBT writing tasks. *Language Testing*, *31*(2), 241-259.

Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviours in independent and integrated speaking tasks. *Applied Linguistics, 34*(3), 304-324.

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2011). *What makes listening difficult? Factors affecting second language listening comprehension.* University of Maryland Center for Advanced Study of Language.

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, *19*(4), 369-394.

Brown, A. (1993). The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated teset of proficiency in spoken Japanese. *Language Testing, 10*, 277-300.

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.

Brunfaut, T., & Révész, A. (2015). The Role of Task and Listener Characteristics in Second Language Listening. *TESOL Quarterly*, *49*(1), 141-168.

Buck, G. (1991). The testing of listening comprehension: an introspective study. *Language Testing, 8*(1), 67-91.

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

Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the input hypothesis. *TESOL Quarterly, 19*(4), 765-781.

Candlin, C. N. (2009). Towards task-based language learning. In K. ven den Branden, B. Martin, & J. M. Norris (Eds.), *Task-based language teaching: a reader* (pp. 21-40). The Netherlands: John Benjamins B.V.

Chalhoub-Deville, M. (1996). Performance and components of the oral content across different tasks and rater groups. In M. Milanovic & N. Saville (Eds.), *Performances, testing, cognition, and assessment* (pp. 55-73). Cambridge: Cambridge University Press.

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

Chapelle, C. A. (2012). Validity argument for language assessment: The framework is simple. *Language Testing*, *29*(1), 19-27.

Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment, 16*(2), 104-122.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Crutcher, R. J. (1994). Telling what we know: the use of verbal report, methodologies in psychological research. *5*(5), 241-244.

Cumming, A. (2014). Assessing integrated skills. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 216-229). Hoboken: John Wiley & Sons, Inc.

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.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL Monograph Series MS30. from Educational Testing Service

Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*

Davies, A. (2005). *A glossary of applied linguistics* Edinburgh: Edinburgh University Press Ltd.

Davies, A. (2012). Kane, validity and soundness. *Language Testing, 29*(1), 37-42.

Deroey, K. L. (2015). Marking importance in lectures: Interactive and textual orientation. *Applied linguistics, 36*(1), 51-72.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press: Oxford.

Douglas, D. (2000). *Assessing languages for specific purposes*. Ernst Klett Sprachen.

East, M. (2012). *Task-based language teaching from the teachers' perspective.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficult of oral proficiency tasks: What does the test-taker have to offer? *Language Testing, 19*(4), 347-368.

Ericsson, K. A., & Simon, H., A. (1993). *Protocol analysis: Verbal reports as data*. USA: Massachusetts Institute of Technology.

Færch, C., & Kasper, G. (1986). The role of comprehension in second language learning. *Applied Linguistics, 7*(3), 257-274.

Ferris, D. &Tagg, T. (1996). Academic listening/speaking tasks for ESL students: Problems, suggestions, and implications. *TESOL Quarterly, 30*(2), 297-320

Field, A. (2005). *Discovering statistics using SPSS (2nd ed.)*. London: Sage.

Field, J. (2012). The cognitive validity of the lecture-based question in the IELTS Listening paper. In L. Taylor & C. Weir (Eds.), *IELTS Collected Paper 2: Research in the reading and listening assessment* (pp. 391-453). Cambridge: Cambridge University Press.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77-151). UK: MPG Books Group.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. New York : 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.

Garrod, S. (1986). Language comprehension in context: a psychological perspective. *Applied Linguistics, 7*(3), 226-238.

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.

Gilakjani, A. P., & Ahmadi, M. R. (2011). A study of factors affecting EFL learners' English listening comprehension and the strategies for improvement. *Journal of Language Teaching and Research, 2*(5), 977-988.

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

Graham, S., Santos, D., & Vanderplank, R. (2008). Listening comprehension and strategy use: A longitudinal exploration. *System, 36*(1), 52-68.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Imhof, M. (1998). What makes a good listener? Listening behavior in instructional settings. *International Journal of Listening, 12*(1), 81-105.

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second language task complexity* (pp. 307-329). Amsterdam: John Benjamins Publishing Company.

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

Kane, M. (1992). Quantitative methods in psychology: An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, USA: Greenwood Publishing Group, Inc.

Kane, M. (2012). Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing, 29*(1), 3-17.

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 5-17.

Kasper, G. (1998). Analyzing verbal protocols. *TESOL Quarterly, 32*(2), 356-362.

Khalifa, H., & Weir, C. J. (2009). *Examining reading*. Cambridge: Cambridge University Press.

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

Lado, R. (1961). *Language testing: the construction and use of foreign language tests*. London: Longman Group Limited.

Lee, Y.-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing, 23*(2), 131-166.

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.

Loftus, G. R., & Loftus, E. F. (1976). *Human memory: The processing of information*. United States of America: John Wiley & Sons, Inc., New York.

Lynch, T. (2009). Teaching second language listening. Oxford: Oxford University Press

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

McNamara, T. F. (1996). *Measuring second language performance*. London: Addison Wesley Longman.

McNamara, T. F. (2000). *Language testing*. Oxford: Oxford University Press.

McNamara, T. F. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly, 3*(1), 31-51.

Messick, S. (1989). Meaning and value in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

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.

Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing*, *19*(4), 395-418.

Norris, J. M., Bygate, M., & Van den Branden, K. (2009). Task-based language assessment. In K. V. d. Branden, M. Bygate, & J. M. Norris (Eds.), *Task-based language teaching* (pp. 431-434). Amsterdam: John Benjamins Publishing Company.

O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics, 10*(4), 418-437.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.

O'Sullivan, B. (2012). Assessment issues in language for specific purpose. *The Modern Language Journal*, *96*, 71-88

Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing, 25*(2), 237-272.

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

Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes, 8*(4), 252-266.

Plakans, L. (2015). Integrated second language writing assessment: why? what? how? *Language and Linguistics Compass, 9*(4), 159-167.

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

Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling appraoch*. Cambridge: The Press Syndicate of the University of Cambridge.

Ren, W. (2013). A longitudinal investigation into L2 learners' cognitive processes during study abroad. *Applied Linguistics*, 1-21.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition, 35*, 31-65.

Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching & applied linguistics*. London: Pearson Education Limited.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27-57.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and Motion: Measuring the Effects of the Conceptual Demands of Tasks on Second Language Speech Production. *Applied Linguistics, 30*(4), 533-554.

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

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Great Britain: Pearson Education Limited.

Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied Linguistics, 11*(2), 117-131.

Sawaki, Y., Quinlan, T., & Lee, Y.-W. (2013). Understanding learner stregths and weaknesses: Assessing performance on an integrateg writing task. *Language Assessment Quarterly*, 73-95.

Sawaki, Y., Stricker, L. J., & Oranjie, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing, 26*(1), 5-30.

Shang, H.-F. (2008). Listening strategy use and linguistic patterns in listening comprehension by EFL learners. *International Journal of Listening*, *22*(1), 29-45

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.

Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: the effect of text and questions type. *Language Testing, 8*(1), 23-40.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics, 17*(1), 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Song, X., & Cheng, L. (2006). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly: An International Journal, 3*(3), 243-266.

Storey, P. (1997). Examining the test-taking process: a cognitive perspective on the discourse close test. *Language Testing, 14*(2), 214-231.

Swain, M., Huang, L.-S., Barkaoui, K., Brooks, L., & Lapkin, S. (2009). *The speaking section of the TOEFL iBT™ (SSTiBT): Test-takers' reported strategic behaviours (TOEFL iBT™ research report)*. Retrieved from Princeton, NJ:

Tavakoli, P. (2009). Investigating task difficulty: learners' and teachers' perceptions. *International Journal of Applied Linguistics, 19*(1), 1-25.

Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes, 10*(2), 89-101.

Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator or L2 listening performance. *Applied Linguistics, 19*(4), 432-451

Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign language Annals, 30*(3), 388-409.

Vandergrift, L. (2003). Orchestrating strategy use: toward a models of the skilled second language listener. *Language Learning, 53*(3), 463-496.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening*. New York: Routedge.

Weigle, S., Yang, W., & Montee, M. (2013). Exploring reading processes in an academic reading test using short-answer questions. *Language Assessment Quarterly, 10*(1), 28-48.

Weigle, S. C. (2004). Integrated reading and writing in a competency test for non-native speakers of English. *Assessment Writing, 9*, 27-55.

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.

Weir, C. J. (2005). *Language teasint and validation: An evidence-based approach*. Great Britain: Antony Rowe Ttd,.

Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education*. Language testing and Assessment, pp. 177-196).

Xie, Q. (2011). Is test-taker perception of assessment related to construct validity? *Internation Journal of Testing, 11*(4), 324-348.

Yi'an, W. (1998). What do tests of listening comprehension test? - A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing, 15*(1), 21-44.

Yu, G. (2013). The use of summaization tasks: Some lexical and conceptual analyses. *Language Assessment Quarterly, 10*(1), 96-109.

Appendices

Appendix 1: Ethics documentation

Date: 24 January 2013

INFORMATION SHEET

As part of my doctoral studies in the Department of Linguistics and English Language, I have been asked to carry out a study involving giving an English language test to students who are not native speakers of English. In my study, I will look into students' test performance and thinking processes whilst doing the test, as well as their perceptions of the test and its difficulty.

I have approached you because I'm interested to know how Thai students, who speak English as a foreign language and need English for higher education, complete this test and how they perceive the test and its difficulty. I would be very grateful if you would agree to take part.

By participating in this study, you will first be asked to provide information on your educational background by filling in a one-page questionnaire. After that, the English language test will be delivered and you will be required to complete 4 tasks. Each task begins with listening to an audio passage and then giving a summary of the passage in either oral or written form. After completing each task, you will be shown the task and your response again and invited to talk about how you approached the task and what you were thinking about while doing it. Video recording devices will be used during this process. Also, I will ask you to fill out a questionnaire in order to know what your views on the test and how difficult it was.

You are free to withdraw from the study at any time. At every stage, your name will remain confidential. Your real name will not be used but assigned a pseudonym. The data will be kept securely in a locked cupboard and electronic data will be saved on a computer protected by password access. The data will be used for academic purposes only.

If you have any queries about the study, please feel free to contact myself, or my supervisor, Dr. Tineke Brunfaut, who can be contacted at t.brunfaut@lancaster.ac.uk or by phone on +44(0)1524 594084. You may also contact the Head of Department, Prof. Elena Semino, at e.semino@lancaster.ac.uk or by phone on +44(0)1524 594176.

Ms. Anchana Rukthong
a.rukthong@lancaster.ac.uk
PhD student

Lancaster University
Lancaster LA1 4YL
United Kingdom
Tel: +44 (0)1524 593045
Fax: +44 (0)1524 843085
http://www.ling.lancs.ac.uk

UNIVERSITY OF LANCASTER

Department of Linguistics and English Language

Consent Form

Project title: A construct validation study of integrated listening-speaking and listening-writing test items

i. I have read and had explained to me by Ms. Anchana Rukthong the Information Sheet relating to this project.
ii. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.
iii. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.
iv. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:____

Signed:_______________________________________

Date:____

Appendix 2: Background Questionnaire

Directions: Please provide the following information by writing your response in the space or ticking (✓) in the box that is true to your information.

1. **First Name**: ___________________________ 2. **Last Name**: __________________________

3. **Gender:** ☐ Male ☐ Female 4. **Age:** ___________ years

5. **1st language**: ☐ Thai ☐ Other: ______________________

6. **Current level of study:**

☐ Undergraduate Year of study: ☐ 1st ☐ 2nd ☐ 3rd ☐ 4th ☐ Other: ___

☐ Masters Year of study: ☐ 1st ☐2nd ☐ Other: _____

☐ PhD Year of study: ☐ 1st ☐2nd ☐ Other: _____

☐ Other: ______________________

7. **Faculty:** ________________________________ 8. **Major subject:** __________________________

9. **Overseas experience:** Have you spent a long period (at least a total of three months) in English speaking countries? ☐ Yes ☐ No

If yes, which country? ______________________________

How long did you live there? _________year(s)_________month(s)

10. **English ability:** Please provide your English language scores.

☐ IELTS ☐ TOEFL

Overall score _________________

Score on each skill: Listening ________ Reading ______Writing ________ Speaking _______

When did you take the test?

☐ Within 3 months ago ☐ Within 3-6 months ago

☐ Within 6-12 months ago ☐ More than one year ago

Thank you very much!

Appendix 3: Listening-to-speak task perception questionnaire

Directions: For each of the following statements, please put a tick (✓) in the column that best represents your level of agreement.

	Statements	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
1	The task simulates situations in academic contexts.					
2	The task assesses the English ability required for academic study.					
3	This is a fair way to assess my ability to use English in academic contexts.					
4	The task accurately reflects my English listening ability.					
5	The task accurately reflects my English speaking ability.					
6	Vocabulary in the listening passage was difficult for me.					
7	Sentence structures in the listening passage were complicated for me.					
8	There were a lot of important ideas to be processed during the listening passage.					
9	Important ideas in the listening passage were paraphrased or repeated more than once.					
10	Ideas in the listening passage were clearly connected.					
11	I am familiar with the content of the listening passage.					
12	I could predict the rest of listening content after listening to the first few sentences.					
13	I have attended academic lectures in English before.					
14	I have the experience of listening to an academic text and then orally summarizing it.					
15	The ideas in the listening passage were organized clearly.					
16	I had to pay attention to more than one idea at a time.					
17	The listening passage provided sufficient ideas for me to complete an oral summary.					
18	The listening passage contained a lot of implied meanings.					
19	The listening passage contained a lot of abstract ideas.					
20	I felt under time pressure while performing the task.					
21	I had enough time to perform the speaking task.					
22	The listening was too long.					
23	The passage was spoken too fast.					

Appendix 4: Listening-to-write task perception questionnaire

Directions: For each of the following statements, please put a tick (✓) in the column that best represents your level of agreement.

	Statements	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	The task simulates situations in academic contexts.					
2	The task assesses the English ability required for academic study.					
3	This is a fair way to assess my ability to use English in academic contexts.					
4	The task accurately reflects my English listening ability.					
5	The task accurately reflects my English writing ability.					
6	Vocabulary in the listening passage was difficult for me.					
7	Sentence structures in the listening passage were complicated for me.					
8	There were a lot of important ideas to be processed during the listening passage.					
9	Important ideas in the listening passage were paraphrased or repeated more than once.					
10	Ideas in the listening passage were clearly connected.					
11	I am familiar with the content of the listening passage.					
12	I could predict the rest of listening content after listening to the first few sentences.					
13	I have attended academic lectures in English before.					
14	I have the experience of listening to an academic text and then writing its summary.					
15	The ideas in the listening passage were organized clearly.					
16	I had to pay attention to more than one idea at a time.					
17	The listening passage provided sufficient ideas for me to complete a written summary.					
18	The listening passage contained a lot of implied meanings.					
19	The listening passage contained a lot of abstract ideas.					
20	I felt under time pressure while performing the task.					
21	I had enough time to perform the writing task.					
22	The listening was too long.					
23	The passage was spoken too fast.					

Appendix 5: Coding scheme

Cognitive process	Strategies	
	Cognitive	Metacognitive
1) Acoustic-phonetic decoding	1) Fixation	1) Pre-listening preparation
2) Word Recognizing	2) Inferencing	2) Selective attention
3) Parsing	3) Elaboration	3) Directed attention
4) Semantic processing at a local level	4) Prediction	4) Comprehension monitoring
5) Semantic processing at a global (discourse) level	5) Translation	5) Real-time assessing of input
6) Pragmatic processing	6) Reconstruction	6) Comprehension evaluation

Cognitive Processes

1) **Acoustic-phonetic decoding**

Occurring when a listener accesses acoustic sounds, registers the sounds, and converts the sounds into the representations of the language phonological system At this stage of processing, phonemes or phonological forms which are the basic units of words are identified.

2) Word recognizing

The process by which the listener segments a continuous speech to identify words or a series of words (phrase) in a speech stream

3) Parsing

Occurring when the listener combines words and maps them onto the syntactic or semantic structures of the language The result of a parsing process is propositions which generally consist of one predicate and one argument (an agent, an object, or a verb modifier) (Anderson, 1985; Kintsch & van Dijk, 1978).

**Semantic processing takes place when the listen relates the text information together and to their world knowledge to understand the text meaning. Semantic processing occurs at two different levels, i.e. a local level and a global level.*

4) Semantic processing at the local level

Occurring when the listener makes a connection between individual propositions

5) Semantic processing at the global level

Occurring when the listener links connected propositions to the theme of a topic This is to conceptualize the text meaning and understand the discourse meaning of the text.

6) Pragmatic processing

Occurring when the listeners try to understand the intended meaning of a text left unsaid/ inexplicitly stated by the speaker, by using pragmatic knowledge to determine speaker's intentions. That is, they elaborate on the linguistic information and their social and cultural knowledge about the context of communication.

Strategies

Cognitive strategies

1) Fixation: focussing one's attention on understanding a small part of a spoken text:

- stop to think about the spelling of unfamiliar words
- stop to think about the meaning of words or parts of the input
- memorise/repeat the sounds of unfamiliar words
- memorise words or phrases for later processing

2) Inferencing: using information within the text or conversational context to guess the meaning of unfamiliar language items associated with a listening task, or to fill in missing information or listening gaps:

- using known words in an utterance to guess the meaning of unknown words
- using tone of voice and/or para-linguistics to guess the meaning of unknown words in an utterance
- using background sounds and relationships between speakers in an oral text, images/ visual cues provided, or concrete situational referents to guess the meaning of unknown words

- using information beyond the local sentential level to guess the meaning

3) Elaboration: using prior knowledge from outside the text or conversational context and relating it to knowledge gained from the text in order to fill in missing information

- referring to prior experience personally
- using knowledge gained from experience in the world
- using the knowledge of the target language

4) Prediction: anticipating contents before and during listening

- anticipating general contents (global)
- anticipating details while listening (local)

5) Translation: changing words, phrases or sentences into L1 before interpretation

6) Reconstruction: using key words to recreate meaning

- reconstruct meaning from words heard and notes taken

Metacognitive strategies

1) Pre-listening preparation: preparing mentally and emotionally for a listening task

- bring to consciousness knowledge of the topic and any relevant cultural information
- prepare the conditions for listening by clearing minds of distractions and focusing attention
- determine where to pay attention and decide on how much detail to find, based on purpose for listening in order to direct listening efforts
- anticipate words or ideas that one may hear

2) Selective attention: noticing specific aspects of input:

- listen to words in group
- listen for gist
- listen for familiar content words

- notice how information is structured (e.g., discourse markers)

- pay attention to repetitions

- notice intonation features (e.g., fall and rise tones)

- listen to specific parts of the input

- pay attention to visuals and body language

3) Directed attention: monitoring attention and avoiding distractions

- concentrate hard

- continue to listen in spite of difficulty

4) Comprehension monitoring: checking and confirming understanding while listening

- confirm that comprehension has taken place

- identify words or ideas not understood

- check current interpretation with context of the message/ with prior knowledge

- check for consistency with their predictions, for appropriateness with word knowledge and for internal consistency: that is, the ongoing interpretation of the co-text

- verify predictions and accept the fact that they do not need to understand every word

5) Real-time assessing of input: making on-the-spot decisions about the value of specific parts of the input

- assess the potential value of unfamiliar words

- determine the importance of subsequent parts of input

6) Comprehension evaluation: checking final interpretation for accuracy, completeness, and acceptability

Appendix 6: Table for data coding

Data	Types of processes/strategies
e.g., ถ: ข้อนี้มีวิธีการฟังอย่างไรบ้างคะ ดูวิดีโอนะคะแล้วช่วยเล่าให้ฟังเลยค่ะฟังอย่างไรบ้าง *Q: How did you listen in this item? Let's watch the video and please explain to me how you listened to it* ต: มันก็ยากเป็นคำศัพท์เศรฐศาสตร์ แล้วยังมีกราฟให้ดูมันก็ยากศัพท์ก็ยาก แล้วสรุปจะต้องดูกราฟ หรือต้องฟังคิดว่าไม่มีรูปดีกว่ามันจะได้โฟกัสทเดียว *A: It (the listening passage) was difficult. There are technical terms in economy. There was a graph. Difficult vocabulary. I was thinking if I had to look at the graph or just listen and ignore the graph. It could be very confusing to do both.*	
A: เอาเป็นว่าตอนแรกที่เขาเริ่มขึ้นมาสมาธิหลุดตื่นเต้น หลุดแล้วก็คิดว่าอะไรวะ แล้วรู้สึกว่าฟังไม่ค่อยเข้าใจได้ยินไม่ clear แล้วพอหลุดแล้วหลุดเลย ก็พยายามกลับมาโดยพยายามฟังคำศัพท์ก็หาคำที่เรารู้จักก็คำที่เราได้ยินบ่อยๆๆ *When it (the listening) stated, I was nervous and was not focused. I was thinking what? I did not understand what I was listening to. It was not clear. I missed this part (of the text). I was trying to direct my attention back by listening for words I was familiar with.*	**Directed attention**
ได้ยินคำว่าcorruption, one trillionซึ่งเราก็รู้ว่าน่าจะเป็นnumber ของอะไรสักอย่างหนึ่งแล้วก็มาลองดูกับกราฟพอมาเริ่มดูกับกราฟก็ไม่เห็นคำว่า one trillion แต่ก็มาเจอคำว่า corrupt payment ได้ยินคำว่า development assistance และก็คำว่าfinancial flows *I heard 'corruption and one trillion'. I knew they described number of something. Then I looked at the graph, I did see 'one trillion', but I saw 'corrupt payment'. I heard 'development assistance' and then 'financial flows'.*	**Word recognizing**
ซึ่งก็พยายามจะlink ว่าไอ้สองกราฟตัวนี้มันเกี่ยวยังไงกับone trillion แล้วพอlink ไม่ได้ว่ามันเกี่ยวยังไงก็เริ่มงง ก็เลยคิดเอาจาก background และสิ่งที่ได้ยินแล้ว ว่ามันน่าจะเป็นว่าเงิน corruption มันประมาณ one trillion *I was trying to link how the two bars (in the graph) were related to one trillion. I could not like them together. I was confused. I inferred from my background and what I heard that corrupted money was about one trillion.*	**Inferencing**

Appendix 7: Examples of data

P1: Listening-to-speak: *Corruption*

Stimulated recall transcript

ถ: เป็นไงมั้ง คะ ข้อสอบ

Q: How was this item?

ต: มันก็ยากเป็นเศรฐศาสตร์ ไม่รู้เรื่องไม่มีความรู้เลย แล้วยังมี กราฟให้ดูมันก็ยากศัพท์ก็ยากแล้ว สรุปจะต้องดูกราฟหรือต้องฟัง คิดว่าไม่มีรูปดีกว่ามันจะได้โฟกัสที่เดียว

A: I think it was difficult. It was all about economy. I did not understand it. I had no background. There was a graph. It made the listening difficult, I think. I was not sure whether I should focus on the graph or the listening text.

ถ: โอเคค่ะ เดียวดูวีดิโอนะคะ แล้วรบกวนช่วยเล่าให้ฟังหน่อยว่าเมื่อกี้ตอนที่ฟังโฟกัสอะไรบ้างคะ ให้ความสำคัญกับอะไรบ้าง

Q: OK. I will play the video recoded during your task performance and can you please explain to me what you were thinking about or paying attention to while you were listening?

ต: เอาเป็นว่าตอนแรกที่เขาเริ่มขึ้นมาสมาธิหลุดตื่นเต้น หลุดแล้วอะไร วะ แล้วรู้สึกว่าฟังไม่ค่อยเข้าใจ ได้ยินไม่ clear แล้วพอหลุดแล้วหลุดเลย….. ตอนนี้ก็พยายามกลับมาโดยพยายามฟังคำศัพท์ ก็หาคำที่เรารู้จัก ก็คำที่เราได้ยินบ่อยๆๆ ซึ่งก็เป็นคำว่า corruption, one trillion, ten times ซึ่งเราก็รู้ว่าน่าจะเป็น number อะไรสักอย่างหนึ่งแล้วก็มาลองดูกับกราฟพอมาเริ่มดูกับกราฟ ก็ไม่เห็นคำว่า one trillion แต่ก็มาเจอคำว่า corrupt payment กับ development assistance และก็มาเจอคำว่า financial flows ซึ่งก็พยายามจะ link ว่า ไอ้สองกราฟตัวนี้มันเกี่ยวยังไงกับ one trillion แล้วพอ link ไม่ได้ว่ามันเกี่ยวยังไง ….ก็เริ่มงง ไม่มี background …. ตอนนี้พยายามจดสิ่งที่ได้ยินแล้ว เอามันหมดแล้วหรอ มันหมดเวลาแแล้ว มันเวลาต้องพูดก็เลยต้องดูที่จดรายละเอียดเกี่ยวกับกราฟไป แทนที่จะพูดว่า lecturer พูดเกี่ยวกับอะไร ก็กลายเป็นว่าหนูมาพูดเกี่ยวกับ graph แทน เพราะว่า หนูฟัง lecturer ไม่เข้าใจ เป็นการ summarize ภาพไป

A: OK, when it started, I was so nervous. I did not understand it (the beginning part of the listening text). I ignored it. I was not clear what it was about. ….Here, I was trying to get my attention back to the listening text, trying to recognize words; the words I was familiar with. I heard 'corruption one trillion'. I knew it was about the number of something. I was trying to catch numbers used to describe the graph. I got 'one trillion', ten times'**.** I looked at the graph. I did not see 'one trillion' but I saw 'corrupt payment' and 'development assistance' and then I saw 'financial flows'. I was trying to link them, thinking how the two bars are related to each other and how they are associated with 'one trillion'. I could link them together. ….I was confused here. I don't have background. …At this moment, I was writing what I heard. I was a bit shocked. It (the audio) was finished. It was about time to orally summarize it. . I looked at my written notes about the graph. Instead of summarizing what the lecturer was talking about, I think I was describing the graph. This was because I did not understand the lecture. It was a summary of the picture.

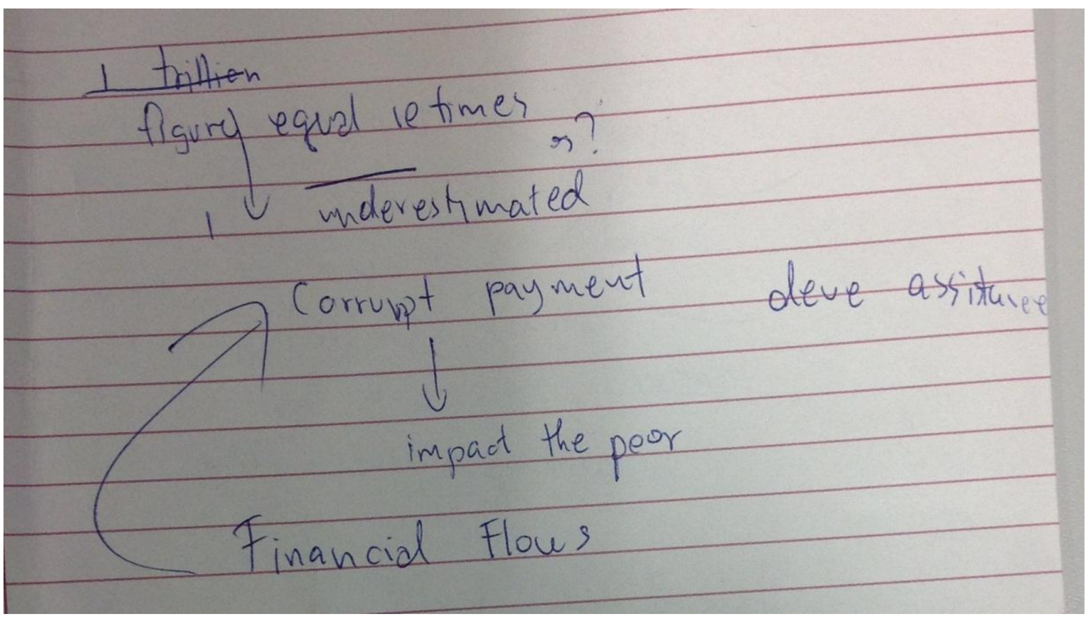
ถ: เดี๋ยวลองฟังอีกรอบนะคะ รอบนี้พี่จะหยุดวีดีโอเป็นช่วงๆ ถ้าระหว่างนั้น มีอะไรจะบอกหรือพูดเพิ่มเติม บอกได้เลยค่ะ

Q: Now, I will play the video again and I will pause it at some points. If you have anything else to explain or tell me, please do so.

ต: ก็ ได้ยินคำว่า corrupt แล้วก็ดูเออว่ามันตรงกับตรงนี้ แล้วเขาพูดประมาณว่า corrupt paymentมันมี impact ต่อ the poor ก็เห็นว่ามันตรงกัน ก็เลยคิดว่า lecture มีอะไร ที่มันน่าจะตรงกัน ก็เลยคิดว่าน่าจะเป็น corrupt ที่มันเกี่ยวกับ the poor ตอนนี้กำลังจดกราฟอยู่เพราะฟังไม่รู้เรื่อง ตรงนี้ฟังเสียงด้วยแล้วก็จดกราฟด้วยแล้วมันก็งงอ่ะ เพราะมีความรู้สึกว่าฟังไม่รู้เรื่องต้องเอาสิ่งที่อยู่ตรงหน้าก่อน เพราะ กลัวว่าจะไม่มีไรพูด ตอนนี้เริ่มเครียด เพราะ ฟังไม่รู้เรื่อง

A: I heard ‘corrupt’. I looked at the graph. It was related to each other. The speaker said ‘corrupt payment has an impact on the poor’. I knew it was related. I assumed that the lecture was about ‘corruption which is related to the poor’….I was noting the graph information because I did not understand the lecture…….I was confused here. I thought I had to pay attention to what I knew and had in front of me so that I had something to talk about in the speaking part. I was worried. I did not understand the lecture.

Note



Oral summary

Ah ..Ah..I think the talk is about the finance flows and according to the chart .. the bar chart ah.... one is about the corruption.. the corrupt payments and the other one is about develop assistance and the speaker talks about a figure something one trillions that the figure like... can impact the poor ..um and.. it's often underestimated. Yeah... and that's all about the talk, I think.

P2: Listening-to-write: *Talent*

Stimulated recall transcript

ถ: ข้อนี้เป็นยังไงบ้างคะ

Q: How was this item?

ต: ข้อนี้มันฟังดู เป็น chronological order คือมันเรียงเนื้อหาในการพูดค่อนข้างชัด มันเข้าใจได้ ง่ายกว่า แค่ มันง่ายกว่าข้อแรก มัน focus ที่เดียวคือ จับ concept ของคำว่า talent

A: There was a chronological order of events. The textual organization was clear so it was easier (than *Corruption*). There was only one focus which is the concept of content.

ถ: ดูวีดิโอนะคะ ตอนฟัง focus อะไรบ้างคะ

Q: Let's watch the video recording. Can you please tell me what you focused on while listening?

ต: ได้ ยินคำว่าtalent แล้วก็คำอื่นที่มีความสำคัญกับนักธุกิจ ..เมื่อกี้ได้ฟังว่าผู้พูดได้พูดถึงความหมายของคำว่า ความหมายที่กว้างที่สุดของคำว่า talent คือ high level analytical ability จะเขียนว่าไงดีว่ะ มันเป็น เอ่อ แล้ว ก็ analytical ability แล้วเขาบอกต่อว่า เขาบอกว่า เพราะมันสำคัญมากเลยสำหรับเศรษฐกิจในปัจจุบัน อันนี้จับไม่ทัน จับไม่ได้ว่ามันคืออะไร she tried to give her own definition

A: I heard 'talent' and other words which were related to business. I heard that the speaker talked about a broad definition of talent: talent is high level analytical ability. ………..I heard analytical ability …The speaker said it (talent) is important to current economy …….I did not know what it was about here. ………I thought ………she was trying to give her own definition.

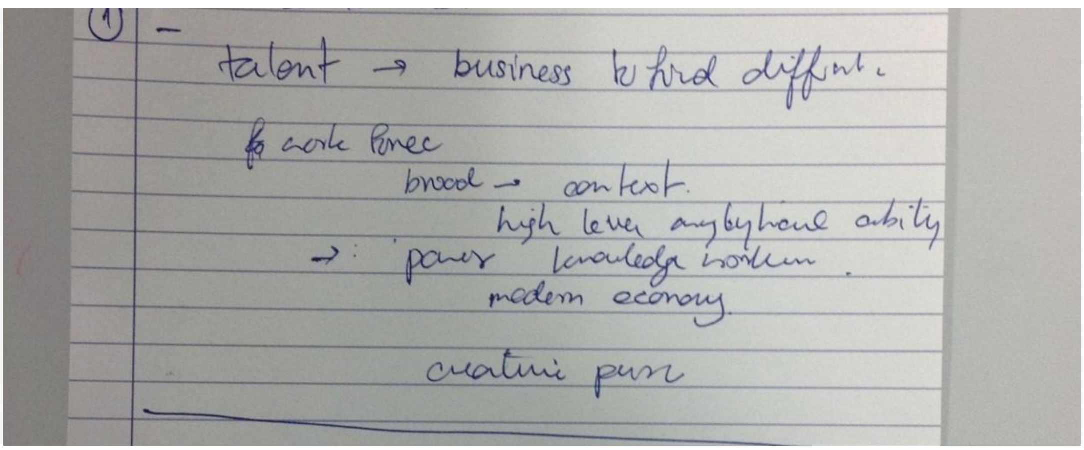
ถ: ดูอีกรอบนะคะ เผื่อมีอะไร เพิ่มเติม

Q: Please watch the video again. I will stop it at different points. If you have something to explain more, please say it.

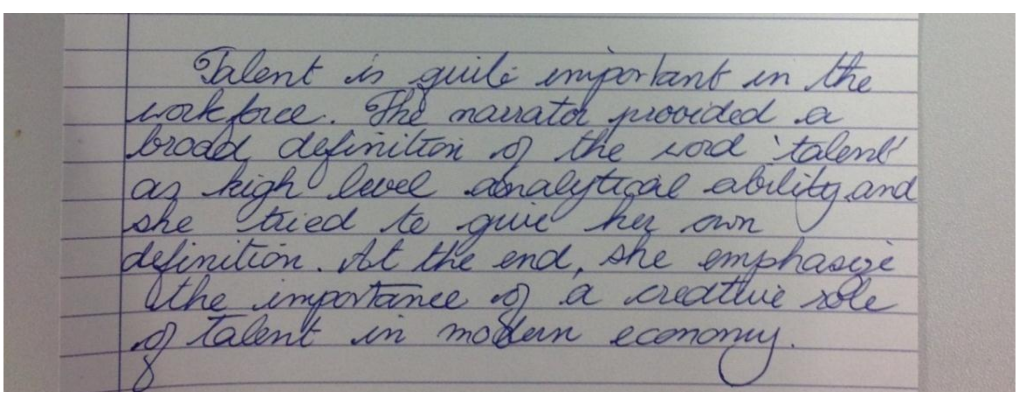
ต: อันนี้ก็ยากครับเพราะมันเป็นเรื่องไกลตัวอีกแล้ว ได้ยิน talent กับ workforce กับ economy ผู้เล่าพูดถึงคำนิยามกว้างๆๆ ของ talent อันนี้ผมจับได้แล้วก็กำลังคิดตามอยู่ อย่างเขาพูดถึงคำนิยามของ talent ต่อไปเรื่อยๆ ก็คือเป็นคำนิยามที่พูดต่อได้เรื่อยๆๆ แต่ ผมจด talent, difficult to define, different view คือว่าผมคิดตามไม่ทันเพราะผมมัวคิดถึงแต่อันแรกอยู่ ยังจดไม่เสด ก็เลย miss ตรงนั้น ไปแล้วก็ miss …..ได้ยินคำว่า workforce ขึ้นมาก็เลยเดาแล้วว่า talent คงมีความสำคัญกับ workforce นะ แล้วก็พูดต่อไปอีกนิดนึง ก็ได้ยินคำว่า modern economy เราก็คิดแล้วว่า ต่อไปว่าก็เลยคิดว่า talent ก็คงสำคัญกับ modern economy ด้วย แล้วก็ฟังต่อไปอีกสักนึดหนึ่งได้ยินคำว่า creative อะไรสักอย่างก็ไม่รู้ ก็เลยเดาไปว่า creative role ของ talent ที่มีต่อ modern economy ตรงนี้ก็ จดๆๆ คำศัพท์ อยู่ แล้ว เงยหน้า มองจอแบบงง เพราะ กำลังคิดว่า เฮ้ยมันพูดถึงไร ว่ะ คืออันนี้ ฟังไม่ทันแล้ว คือจิงๆๆ ก็ฟังอยู่ และพยายามจะจับประเด็น แล้ว ก็งง ว่าประเด็นมันคืออะไร ฟังไม่ทันแล้วก็ ไปต่อไม่ถูกแล้ว ก็เลยเงยหน้าคิดมา ตั้งสติ หยุดเขียน ดูว่ามันจะพูดอะไรต่อไป ตรงนี้ผมงง .... มันอะไรสักอย่าง brib.. power คือผมไม่รู้ว่ามันคืออะไร

A: It was difficult. It was far from my background. I heard ‘talent’, ‘workforce’, and ‘economy’. I knew that the speaker was talking about a broad definition of talent. I was following him…...He (the speaker) continued talking…… I noted down ‘talent’, ‘difficult to define’, and ‘different views'…… but here I could not process the text because I was thinking about the previous idea. I did not finished my note taking….I missed this part…um...I heard ‘workforce’. I guessed talent is important for workforce…he continued talking ….I heard modern economy. ….I predicted that talent is important for modern economy, too. I continued listening …I heard ‘creative and something I did not know. So I guessed it was a creative role of talent in modern economy…..I was writing vocabulary here. I was confused. ….then I looked at the screen, trying to understand the point…I could not catch it. I continued listening, but I did not understand. I stopped writing and listened, trying to figure it out…I was confused here. It was about brib….power something.

Note


(1) –
talent → business to find diffint.
fo work force
broad → context.
high level anylytical ability
→ power knowledge worker.
modern economy
creative pur

Written summary


Talent is quite important in the workforce. The narrator provided a broad definition of the word 'talent' as high level analytical ability and she tried to give her own definition. At the end, she emphasize the importance of a creative role of talent in modern economy.