

**DEVELOPING TEST SPECIFICATIONS FOR
MEASURING LISTENING PERFORMANCE OF
TERTIARY STUDENTS**

IRMA BINTI AHMAD

UNIVERSITI SAINS MALAYSIA

2023

**DEVELOPING TEST SPECIFICATIONS FOR
MEASURING LISTENING PERFORMANCE OF
TERTIARY STUDENTS**

by

IRMA BINTI AHMAD

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

May 2023

ACKNOWLEDGEMENT

(In the name of Allah, the most Gracious and the most Merciful)

Now is the time to thank those who provided me the with opportunity to do this Ph.D supported, and believed in me. First, I would like to thank the Department of Civil Service Malaysia, who awarded me the scholarship, and my employer Universiti Teknologi MARA, who approve a three-year leave to complete the Ph.D. My biggest and deepest appreciation goes to my supervisors. Assoc.Prof. Dr Mohamad Jafre Zainol Abidin, and Dr Al Amin who has been very patient and understanding in guiding me throughout this tedious journey. I would not have made it without their guidance and support in spite of their tight schedule. I would also like to express my sincere appreciation to my colleagues Miss Afni Anida Adnan and Dr. Tuan Sarifah Aini Syed Ahmad, for their tremendous support and commitment in helping me through difficult times and for their encouraging support and assistance. In the process of struggling through this thesis, I would like to express my deepest gratitude, especially to my loving husband, Mohd Shahrul bin Ibrahim, my parents Ahmad Md Isa and Rusni Ismail, and the rest of my family members who have shown me their deepest love and understanding all these years that have indeed helped me in completing this thesis. To complete this thesis was by no means easy. This endeavour has given me vast knowledge and the inspiration to continue working, but most of all, it has added special meaning to my life, and for this, I thank you all. To all of you, I dedicated this thesis.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xiii
LIST OF APPENDICES	xiv
ABSTRAK	xv
ABSTRACT	xvii
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Background of the Study	3
1.3 Problem Statement	5
1.4 Objectives of the Study	8
1.4.1 Objective 1	8
1.4.2 Objective 2	8
1.5 Research Questions	9
1.6 Significance of the Study	10
1.7 Limitation of the Study	11
1.8 Operational Definition.....	12
1.9 Summary of the Chapter	16
CHAPTER 2 LITERATURE REVIEW	17
2.1 Introduction	17
2.2 Listening Comprehension	17
2.3 Overview on English Language Teaching in Higher Education Malaysia	22
2.3.1 The Malaysian University English Test (MUET).....	23

2.3.2	Common European Framework for Languages: Learning, Teaching, Assessment (CEFR)	30
2.3.3	Issues and Challenges of ELT in Higher Education Malaysia.....	35
2.4	The Significance of Listening in Learning.....	39
2.5	Listening Test Specification.....	42
2.6	Response Format	50
2.6.1	Multiple-Choice Question	51
2.6.2	Dictation.....	52
2.6.3	Matching Responses.....	54
2.6.4	True and False	55
2.6.5	Close Passage or Gap-Filling.....	56
2.6.6	Short Answer.....	57
2.6.7	Information Transfer	58
2.7	Listening Skills.....	63
2.7.1	Bloom’s Taxonomy.....	66
2.7.2	CEFR Can-do-Abilities	71
2.8	Validating the Listening Comprehension Test.....	81
2.8.1	Tentative Model of L2 Listening Comprehension Assessment	81
2.8.2	Socio-cognitive Framework for Validating Listening Comprehension Test.....	83
2.9	Theories Underpinning the Listening Test Specification.....	87
2.9.1	Framework of Language Competence	88
2.9.2	Comprehension Approach (CA)	90
2.9.3	Cognitive Processing Model	92
2.10	The Conceptual Framework of the Research	95
2.11	Justification of the selection of the elements for the listening test specification based on learning theories	98
2.12	Summary of the Chapter	99

CHAPTER 3	RESEARCH METHODOLOGY	101
3.1	Introduction	101
3.2	Research Design	101
3.3	Instruments	104
3.3.1	Questionnaire	104
3.3.2	Focus group interview	105
3.3.3	Listening Comprehension Test.....	107
3.4	Validity and Reliability of Instruments	107
3.4.1	Reliability of Listening Comprehension Test	108
3.4.2	Content Validity of Questionnaire	109
3.4.3	Content Validity of Listening Specification (LTS) & Listening Performance Descriptors (LPD).....	111
3.4.4	Validity of Questions in the Focus Group Interview	111
3.5	The Participants of the Study	112
3.6	The Sampling of Research Participants	113
3.7	Research Procedures	114
3.7.1	Test Elements for Listening Comprehension Test	115
3.7.2	Thematic Analysis of Focus Group Interview	116
3.7.3	Data collection procedures	117
3.8	Generating Validity of the Test	118
3.8.1	Context Validity	118
3.8.2	Theory-Based Validity	118
3.8.3	Scoring Validity	119
3.8.4	Consequential Validity & Criterion-Related Validity.....	121
3.9	Summary of the Chapter	122
CHAPTER 4	THE DEVELOPMENT OF LISTENING TEST SPECIFICATION	123
4.1	Introduction	123

4.2	The Listening Test Specification (LTS).....	123
4.3	The Development of the Listening Test Specification.....	124
4.3.1	Generating Context Validity of the Test Specification	125
4.3.2	Generating Theory-Based Validity	128
4.3.3	Generating Scoring Validity	129
4.3.4	Generating Consequential Validity	130
4.4	Generating Criterion-related Validity	131
4.5	Listening Comprehension Test (LCT)	132
4.6	Listening Performance Descriptors	133
4.7	The overall phases of the developing the LTS, LCT and LTD.....	134
4.8	Conclusion.....	140
CHAPTER 5 RESULTS AND FINDINGS		150
5.1	Introduction	150
5.2	Context Validity of the Test	150
5.2.1	Generating Context Validity of the Test	150
5.2.1(a)	Quantitative Analysis.....	152
5.2.1(b)	Qualitative Analysis.....	160
5.2.2	Generating Theory-based Validity of the listening test	169
5.2.3	Generating Scoring Validity	170
5.2.3(a)	Item Analysis and Item Discrimination.....	170
5.2.3(b)	Analysis of the test score	174
5.2.4	Generating Consequential Validity	199
5.2.5	Generating Criterion-related validity	201
5.3	Conclusion.....	203
CHAPTER 6 DISCUSSION AND RECOMMENDATION		204
6.1	Introduction	204
6.2	Discussion of the Research	204

6.2.1	The Development of the Listening Test Specification.....	205
6.2.2	Development of the Listening Comprehension Test.....	206
6.3	Differences between the LTS and other listening test specifications.....	209
6.4	Implications of the Study	211
6.5	Recommendations	214
6.6	Conclusion.....	215
	REFERENCES.....	218

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

	Page
Table 2.1	MUET Listening Report from 2017-2019 (Statistics: Malaysian Examination Council, 2020)24
Table 2.2	The inaccurate attempts made by the MUET candidates in the listening component (Malaysian Examination Council, 2021).....26
Table 2.3	Comparison between the old and the revised MUET listening test specifications as adapted in Malaysian Examination Council (2006) and Malaysian Examination Council (2021).....28
Table 2.4	Hours allocated for English and CEFR Guided Learning Hours (p.327, the English Language Education Reform: Roadmap 2015-2025)34
Table 2.5	Comparison of listening test structure of the older version and the latest version of MUET46
Table 2.6	The revised MUET listening test specifications47
Table 2.7	Band descriptor for the listening component of MUET adapted from (Malaysian Examination Council, MUET: Regulations, Test Specifications, Test Format and Sample Questions, 2021).....49
Table 2.8	Summary of literature search for response format.....59
Table 2.9	Comparison of definition and structure of the original and revised Bloom's Taxonomy (adapted from Krathwohl, 2002)68
Table 2.10	Comparison of verbs for testing between the original and revised Bloom's Taxonomy (Krathwohl, 2002)69
Table 2.11	CEFR Overall Oral Comprehension Level (Council of Europe, 2021)73
Table 2.12	Understanding conversation between other people (Council of Europe, 2021).....74

Table 2.13	Understanding as a member of a live audience (Council of Europe , 2021).....	76
Table 2.14	Understanding announcement and instructions (Council of Europ , 2021)	77
Table 2.15	Understanding audio (or signed) media and recordings (Council of Europe , 2021).....	78
Table 2.16	The key concepts operationalized the fours scale of the CEFR oral comprehension as adapted from (Council of Europe, 2021)	80
Table 2.17	Summary of elements selected for a listening test specification.....	98
Table 3.1	Overall research design of the study	103
Table 3.2	Reliability statistics listening comprehension test of phase three and four of the study	109
Table 3.3	Comments and amendments made for Context Validity Checklist .	109
Table 3.4	Population and sample size of the study	114
Table 4.1	Comparison elements selected for LTS with MUET.....	136
Table 5.1	Context validity for format of the test from students' perspective ...	153
Table 5.2	Context validity for format of the test from experts' perspective	154
Table 5.3	Context validity for format of the test from students' point of view	156
Table 5.4	Comments from the experts and students on the listening test comprehension	157
Table 5.5	The experts' comment on the listening test descriptors.....	159
Table 5.6	Focus group interview finding for the purpose of the test	162
Table 5.7	Focus group interview findings for the response format	164
Table 5.8	Focus group interview findings for order of items.....	167
Table 5.9	Focus group interview findings for content knowledge.....	168
Table 5.10	Focus group interview findings for interlocutor	169
Table 5.11	Standard error measurement for the test in the study.....	172

Table 5.12	Item difficulty and discrimination of pilot test	174
Table 5.13	The score range based on the cut score from the listening test score	174
Table 5.14	Item analysis for Dictation	177
Table 5.15	Example of students' responses	178
Table 5.16	Item analysis for True and False	182
Table 5.17	Item analysis for information transfer items	185
Table 5.18	Item analysis for map and labelling items	187
Table 5.19	Item analysis of Short Answer items	190
Table 5.20	Item analysis for gap filling items.....	193
Table 5.21	Item analysis for matching responses items.....	194
Table 5.22	Item analysis for Information Transfer II.....	196
Table 5.23	Item analysis for MCQ items	198
Table 5.24	Focus group interview findings for the most preference question type.....	200
Table 5.25	Focus group interview findings for the test improvement	200
Table 5.26	Kruskal Wallis test statistic for test score and MLS	202
Table 5.27	Correlation between listening test score and MUET listening score	202

LIST OF FIGURES

	Page
Figure 2. 1	Analysis of results by using keyword "listening" and "Malaysia"36
Figure 2. 2	Analysis of results by using the keywords "MUET" and "Malaysia"37
Figure 2. 3	Reception activities and strategies in the CEFR (Council of Europe, 2021).....72
Figure 2. 4	The tentative model of L2 listening comprehension assessment82
Figure 2. 5	A socio-cognitive framework for validating listening comprehension test (Weir,2005: p.45).....84
Figure 2. 6	Some components of Language Test Performance adapted from Bachman and Palmer (1996: p.63).....89
Figure 2. 7	The Cognitive Processing Model (Geranpayeh & Taylor, 2013)94
Figure 2. 8	The conceptual framework of the study.....97
Figure 3. 1	Research design adapted from Weir's socio-cognitive framework ..102
Figure 3. 2	The Convergent Parallel Mixed Methods Design (Adapted from Clark & Creswell, 2015. p.391)103
Figure 4. 1	The context validity elements by Weir's (2005)125
Figure 4. 2	Theory-based validity from Weir's socio-cognitive framework128
Figure 4. 3	Scoring validity from Weir's socio-cognitive framework.....130
Figure 4. 4	Consequential validity from Weir's socio-cognitive framework131
Figure 4. 5	Criterion-related validity from Weir's socio-cognitive framework..132
Figure 4. 6	The overall phases of the study134
Figure 5. 1	Distribution of pilot study test score171
Figure 5. 2	The cut-off points for band 1-5175

Figure 5. 3 The map of City Shopping District..... 188

LIST OF ABBREVIATIONS

CEFR	Common European Framework of Reference
CLT	Communicative Language Teaching
ELT	English Language Teaching
LCT	Listening Comprehension Test
LTD	Listening Test Descriptors
LTS	Listening Test Specifications
MOE	Ministry of Education
MUET	Malaysian University English Test

LIST OF APPENDICES

Appendix A	Questionnaire 1 (Preliminary Study)
Appendix B	Context Validity Checklist for LCT (Student)
Appendix C	Listening Comprehension Test (LCT)
Appendix D	Answer Key
Appendix E	LCT Typescript
Appendix E	Summary of Item Difficulty
Appendix F	Permission to Collect Data
Appendix G	List of Achievement

PEMBANGUNAN SPESIFIKASI UJIAN UNTUK MENGUKUR PRESTASI MENDENGAR PELAJAR PENGAJIAN TINGGI

ABSTRAK

Sejak 2010, laporan ujian MUET bagi komponen pemahaman pendengaran telah menunjukkan hampir separuh calon yang menduduki ujian tersebut dilaporkan mempunyai kemampuan yang terhad dalam kemahiran mendengar dan memahami di dalam Bahasa Inggeris. Ini menunjukkan terdapat jurang kemahiran mendengar dalam Bahasa Inggeris di antara pelajar pra-universiti dan pelajar universiti. Kajian awal telah dijalankan dan hasil dapatan kajian menunjukkan pelajar tidak mengetahui kemampuan mereka dalam kemahiran mendengar dan bersetuju untuk mengukur kemampuan mendengar mereka dengan satu ujian alternatif. Namun begitu, hasil kajian literatur mendapati terdapat kekurangan garis panduan bagi membina ujian tersebut terutama untuk pelajar peringkat universiti. Oleh yang demikian, kajian ini dijalankan dengan matlamat untuk membangunkan satu spesifikasi ujian kemahiran mendengar di dalam Bahasa Inggeris untuk pelajar di peringkat universiti. Terdapat dua objektif kajian iaitu: (1) untuk membangunkan spesifikasi ujian kemahiran mendengar bagi tujuan pengukuran kemahiran mendengar dikalangan pelajar ijazah sarjana muda, dan (2) menggunakan kerangka ujian dari Weir untuk tujuan validasi spesifikasi ujian yang dibangunkan dalam kajian ini. Borang soal-selidik, temubual, dan borang pengesahan kandungan telah digunakan untuk mengumpul data bagi kajian ini. Manakala untuk menganalisa data dari kaedah kuantitatif, SPSS telah digunakan dan untuk menganalisa temubual, analisis tematik telah digunakan. Hasil dapatan dari kajian ini ialah tujuh elemen untuk spesifikasi ujian iaitu: (1) Spesifikasi Am, (2) struktur ujian, (3) spesifikasi ujian, (4) spesifikasi penyampaian ujian, (5)

spesifikasi markah ujian, (6) tahap kesukaran ujian, (7) penerangan keupayaan ujian kemahiran mendengar. Satu ujian kemahiran mendengar telah dibina dengan menggunakan spesifikasi ujian dan ujian tersebut telah dilaporkan sesuai untuk digunakan bagi menguji tahap kemahiran mendengar pelajar. Selain itu, satu set penerangan keupayaan kemahiran mendengar pelajar yang mengandungi 5 tahap telah dibangunkan dan digunakan untuk intepretasi markah ujian. Dapatan kajian juga telah menunjukkan 28% atau 109 peserta kajian dikategorikan ‘di bawah jangkaan’ pemahaman mendengar yang telah ditetapkan. Tiga bentuk soalan iaitu *Dictation, true and false dan map labelling* telah dilaporkan sebagai mencabar dan disukai oleh peserta kajian. Hasil dapatan kajian dapat dijadikan sebagai garis panduan kepada pengajar Bahasa Inggreris atau penggubal soalan dalam membangunkan soalan untuk ujian kemahiran mendengar khususnya untuk pelajar di peringkat universiti.

DEVELOPING TEST SPECIFICATIONS FOR MEASURING LISTENING PERFORMANCE OF TERTIARY STUDENTS

ABSTRACT

Reports from the Malaysian Examination Council have shown a worrying trend in the Malaysian University English Test (MUET) listening component since 2010, where nearly half of the candidates were reported as ‘limited’ and ‘very limited’ users. This indicates a lack of listening skills among the candidates and thus, proof that there is a gap in English language listening proficiency between pre-university and tertiary education. A preliminary study was conducted in the study and participants were reported as not being not aware of their listening abilities and agreed to have a listening comprehension test to measure their listening proficiency skills. However, there are limited guidelines for designing and developing effective listening comprehension tests, especially for the tertiary students. Thus, this study was aimed to develop a listening test specification that can be used to measure tertiary students’ listening performance. There were two objectives of the study: (1) to develop a listening comprehension test specification to measure students’ listening performance, and (2) to validate the test specification by using elements from Weir’s socio-cognitive framework. The study employed a mixed method. Questionnaires, focus-group interviews, and content validation were used to collect the data. The quantitative data collected were analysed by using Statistical Packages for Social Sciences (SPSS) and thematic analysis was used to analyse the data collected from focus-group interview. The results of the study were the selection of elements for the test specification: (1) general specification, (2) test structure, (3) task specification, (4) delivering specification, (5) scoring specification, (6) test difficulty level, and (7)

test descriptors. A listening comprehension test was developed based on the test specification and found to be valid and reliable to measure participants' listening performance. Dictation, true and false and map labelling were found to be challenging, but they were the preferred task among the participants. Overall, the test specification developed from this study is suitable to measure tertiary students' listening performance. The results of the study can be compiled as a guideline to benefit language instructors and test developers in developing listening comprehension tests, particularly for tertiary students.

CHAPTER 1 INTRODUCTION

1.1 Introduction

Despite its obvious importance in language learning, listening skill has long and widely been acknowledged as a neglected and undervalued skill. With the initiation of communicative language teaching and the emphasis on competency, the teaching and learning of listening started to receive more attention. Learning listening is challenging as it involves multiple settings, various rates of delivery, and a variety of language structures. Unlike reading a text, the process is controlled by the learner. Whereas in listening, learners do not have any control over the text as it constantly moves at variable speeds and sometimes learners need to learn to comprehend the reduced form of the language to understand what the speakers say. Besides teaching and learning, the assessment of listening proficiency is also one of the least understood, least developed and yet, one of the most important areas of language testing and assessment (Khoii & Paydarnia, 2011; Ahmad Razaei, 2018).

Listening lies at the heart of language learning (Field, 2008; Rost, 2014) and assists second language (L2) learning (Buck, 2001; Ockey & Wagner, 2018). Listening is an important skill, especially for tertiary students in Malaysia as most lectures and discussions are conducted using the English language in the country. It is also one of the components in the Malaysian University English Test (MUET), which is a compulsory test for students to take before entering any university in Malaysia. However, after 11 years of learning English (six years in the primary school and five years in the secondary school), students are found to have low English proficiency (Nor Hashimah et al., 2008; Darus & Subramaniam, 2009); even based on MUET reports, the listening component was found to be the lowest component scored by students (Majlis Peperiksaan Malaysia, 2020). However,

teaching and learning listening has been neglected in the Malaysian classroom (Suchitra et al., 2013; Suzanah & Sidhu, 2013; Nur Anneliza et al., 2020).

There is no doubt that listening proficiency tests have existed for many decades and the tests have consequently shifted from the traditional methods of listening comprehension which involves listening to the examiner reading aloud the passages to listening to pre-recorded materials. With the emergence of technology and digital networking, instructors are not limited to offering the students conventional pen-paper listening activities, but also online learning activities that come in many forms of online listening quizzes and online tests. Since there are various forms of testing listening nowadays, issues related to construct validity of the test such as how the test is being developed, who will sit for the test, type of response format to be used for the test, speakers' speed and system to grade the test score; need to be given attention. These issues do not occur only in the local context, such as MUET which is facing construct validation issues (Elia, 2014), they also occurred at international high-stake assessment such as IELTS (Alavi et.al, 2018).

This study focuses on developing a test specification for measuring the listening performance of tertiary students. Thus, this chapter provides information that leads to the understanding of why developing a test specification for measuring the listening performance of tertiary students need to be emphasized, and what to be considered in designing and developing the test specification. To understand why developing listening test specification needs to be done, the current practices of listening skills in the Malaysian classroom and public assessment in Malaysia are discussed. Some issues associated with the overall performance of students in UiTM and performance of the MUET candidates in the listening component are also

presented. The problems of the study, objectives, research questions, significance and limitations are discussed in this chapter.

1.2 Background of the Study

Universiti Teknologi MARA (UiTM) places a great emphasis on ELT. English is the medium of instruction in UiTM for all courses in preparatory, diploma, and degree programmes except for courses from the third modern languages (Kaur & Rohayah, 2006). Students are required to pass at least band 3 of MUET in order to enroll in undergraduate courses in their chosen field of study. It is also a mandatory requirement for graduates to obtain a minimum grade of “C” for all English courses offered at UiTM. Although students are required to sit for the MUET to show that they have an adequate level of English ability to follow the courses offered at UiTM, prior research indicated that UiTM students have a low English language proficiency level. One of the reported reasons is due to a lack of listening skills (Mohd Salleh, et al., 2018; Khalijah et al., 2019). Students at the tertiary level, whether local or abroad, including UiTM students were found to be anxious when listening to the English language (Mohammad Nasim & Zailin Shah, 2014; Salwani et al., 2018; Nur Afiqah, 2015; Chin et al., 2016; Hafezian & Shafiee, 2018).

Despite placing great emphasis on the use of the English language in UiTM, there are some issues in teaching and learning of English language such as inadequate time to teach and to provide sufficient exposure for enhancing the students’ language skills (Mah et al., 2013), and mixed ability of student’s language proficiency level in English classes in UiTM (Dayang Zarinah, 2014). These have caused problems for the class instructors as they have limited time to teach and yet need to apply different teaching methods to accommodate students with different levels of English proficiency in the class.

Besides that, English language courses offered at the diploma level are mainly to empower all the English language skills and prepare the students to sit for the MUET. Diploma students should undergo three semesters of English language courses and pass all of them before graduation. Therefore, most of the English courses are designed with the influence of the MUET syllabus. There are parts of the assessments such as the speaking and listening assessment, that are constructed by considering the MUET format. For example, the type of response formats in the listening test, and discussion format for the speaking test follow the MUET format. With exposure to the MUET format since the diploma level, students should get enough practice and be ready to sit for the MUET as it is compulsory for those who wish to further their studies at the degree level. However, there were studies reported on too much emphasis on the MUET syllabus and listening skills being neglected at the school and diploma level has affected their listening proficiency (Khatijah & Gurnam, 2004; Kaur & Nordin, 2006; Nair, 2010; Nair & Mathai, 2010; Nur Sakinah & Nooreen, 2013; Nurulmatin, 2014). Undergraduates students are found to be lacking in note-taking skills and this has caused a problem for them when listening to lecturers or discussions (Rosniah, 2006).

Students should be exposed to various ways of response format in testing listening, not just as stipulated in the MUET format. Various response formats can ensure different listening skills are being tested (Buck , 2001; Weir, 2005; Geranpayeh & Taylor, 2011). Moreover, students have different learning styles and proficiency levels thus how listening skills are being tested should be varied too. In fact, due to limited vocabulary and low proficiency in the English language, it was found that students did not perform well in their listening tests (Norazean et.al,

2017). Thus, students should be given more exposure to natural, authentic conversations, taught effectively, and assessed accordingly on their listening skills.

With the current practices of teaching listening skills in the classroom, including the issues of stressing too much on the MUET syllabus and assessments, how far the students at UiTM have learned and mastered the listening skills is doubted. As mentioned earlier, listening is important for students at the tertiary level as they need to comprehend most lectures that are conducted in the English language. UiTM is the perfect place to conduct a study on how a listening test specification should be developed and tested so that the test can be used to benchmark the students' listening proficiency level.

1.3 Problem Statement

The listening skills are crucial for tertiary students because, in academic listening, they do not only need to process the linguistic and non-verbal aspects of communication (e.g., facial cues, gestures) in real-time but also engage with other tasks (e.g., reviewing notes in a handout) and distractions (e.g., classmates talking about irrelevant topics). The complexity of the listening process itself and neglecting the listening skills in the classroom have resulted in lack of listening skills among students. A similar task type experienced by the students in learning and testing listening, due to high dependence on the MUET test specification; has limited the chance for the students to learn and apply suitable listening strategies to construct the meaning. At the same time, the test items of the MUET listening component were found to have unclear keys, to be guessable, and to be based on incoherent tests or personal opinions (English Language Standards and Quality Council, 2015).

As mentioned earlier, based on previous studies at UiTM, the current practice of teaching and assessing listening skills is highly dependent on the MUET syllabus

and it has affected the students' listening proficiency. A preliminary study was conducted at UiTM Negeri Sembilan to find out the issues and challenges of listening comprehension among the students. These issues and challenges are including students' perception of the current listening comprehension test, their listening performance, and their anxiety toward listening. The study was conducted on 236 students. Based on the findings, there were several issues reported on their perception of their English language listening skills. From the results, 64.9% of the students felt that listening comprehension in English was a challenging task for them. 75.5% of them were aware of their proficiency level based on their MUET listening score. However, 77.6% of them were not aware of what their strength and weaknesses were, based on the score that they obtained. As for the skills that make listening to English difficult to them, 62.5% found that paraphrasing while listening to the spoken text was difficult, 59.1% mentioned distinguishing the relevant from irrelevant information as difficult, 56.1% stated recognizing supporting details as difficult, and 51.5% mentioned interpreting the speaker's view, attitudes, and intentions as difficult. The findings also revealed 85.5% of them would like to test their listening comprehension skills to benchmark their performance for improvement purposes. As for listening anxiety, the findings revealed that 51.5% of them have a rather high level of listening anxiety and, 36.4% of them were at a moderate level of listening anxiety.

Findings from the focus interview of eight students revealed that all the students agreed that listening comprehension skills were very important once they entered university and wished to master the skills beforehand. As for their experiences in learning listening skills, all of them stated that they rarely learned listening at school. However, they learned it when they were at the diploma,

matriculation program, and at post-secondary school or in form 6. All of them stated that the listening activities that were done were mostly by using MUET course books. As for their experiences in testing listening, all of them mentioned the MUET was the last test that they had encountered before they entered UiTM. Four students stated the number of questions in the MUET test was appropriate however all the students stated that they would like to have more variety in terms of situations and test type. Five of them thought that the MUET listening component needed to offer more variety in terms of test type to suit their needs as university students. Five of them stated that they had been exposed to similar types of questions and task situations when it comes to listening tests and thus, would like to be exposed to other ways too.

Based on the findings of the preliminary study, it can be concluded that listening comprehension is a challenging task for the students at UiTM Negeri Sembilan. Not knowing the strength and weaknesses of their listening skills and anxiety towards certain areas while listening in English could affect their learning process. Exposure to similar types of testing formats does not benefit the students too. Introducing a variety of test formats will benefit the students as they will be exposed to various tasks and situations which will enhance their experience. At the same time, various tasks format can ensure the students utilize both lower- and higher- levels of the cognitive process when listening to the task given and thus can affect the students' listening performance (Buck, 2001; Brindley & Slatyer, 2002; In'nami & Koizumi, 2009; Wei & Ying, 2017).

It is evident that, designing and developing an alternative listening test specification should be considered. Instructors and students need an instrument or a test that can be an alternative to the one that is available in the market to benchmark

the students' listening performance. The listening test specification offers several benefits that can address the challenges of ELT in tertiary education, especially at UiTM, particularly in listening comprehension. In the context of the study, this test specification can provide the instructors with an alternative way to develop a listening comprehension test to measure their students' listening proficiency level. The development of the test specifications considers elements that are lacking in the current test specifications, it can benefit both the instructors and students as they can apply the various skills suggested in the test specification to measure students' listening performance.

1.4 Objectives of the Study

The main objective of the study is to develop a listening comprehension test specification by considering the elements in the Weir's socio-cognitive framework, to measure the listening performance of tertiary students in Malaysia. In order to do that, the following objectives are set.

1.4.1 Objective 1

To develop a listening comprehension test specification to measure students' listening performance.

1.4.2 Objective 2

To validate the listening comprehension test developed based on the listening test specification by using the elements from the Weir's Socio-cognitive framework for validating the listening comprehension test. The elements are:

- a) context validity of the test

- b) theory-based validity
- c) scoring validity
- d) consequential validity
- e) criterion-related validity

1.5 Research Questions

For the first objective, the development of the test specification is discussed in Chapter Four. The research questions for the second objectives of the study are as follows:

How can the listening comprehension test developed based on the listening test specification be evaluated?

- a. What is the context validity of the test?
 - i. What is the task setting of the test?
 - ii. What is the task demand of the test?
- b. What is the theory-based validity of the test?
- c. What is the scoring validity of the test?
 - i. What is the item analysis of the test?
 - ii. What is the internal consistency of the test?
 - iii. What is the marker reliability of the test?
- d. What is the consequential validity of the test?
 - i. How is the score of the test being interpreted for washback purposes?
- e. What is the criterion-related validity of the test?
 - i. What is the score value of the test?

1.6 Significance of the Study

This study serves as a guideline for instructors who are responsible for drawing up tests of listening ability and for other professionals who may not be actively involved in teaching but have some need to construct or evaluate a listening test, or to use the information that such test provides. This study is significant from the theoretical and measurement perspectives. Hughes (1989) makes a convincing remark on the importance of construct validation in relation to its theoretical implication that is “Construct validation is a research activity, the means by which theories are put to the test and are confirmed, modified, or abandoned. It is through “construct validation that language testing can be put on a sounder, more scientific footing” (p.27). This means by validating the items in the test specification and listening comprehension test of this study, the underlying theories such as cognitive processing and socio-cognitive framework are put under scrutiny and investigated to determine whether the theory matches practices. Therefore, verifying the construct of the listening comprehension test could ascertain whether the test really measures what it is intended to do and students’ listening performance could be determined.

Instructors who may have to design a listening test to measure students’ listening performance can benefit from the listening test specification developed in the study as it can be a guideline to develop an alternative test to measure listening proficiency levels. Results from the test score could be used by the instructors to identify the students’ abilities. As the test specification developed in this study includes the elements that are lacking in other listening test specification that are available in the market, more listening abilities could be tested by using various task formats and setting. Besides that, instructors can use the listening performance descriptors developed in this test to identify their students’ listening performance.

From the scores and the level possessed by the students, instructors will be able to identify which students are lacking in listening abilities and it will facilitate for instructors to guide them during listening activities in the classroom. Instructors can work on areas in which the students need scaffolding. In addition, the information gained on the students' proficiency could help the instructors to improve their teaching instructions by amending or revising their strategies in lecturing. This includes tailoring selections of vocabulary and phrases and speech rates to ensemble their students' capabilities. As for the students, they are able to identify their listening abilities based on the test score and its interpretation. From the score interpretation, immediate measures can be taken by them to enhance the specific area of listening comprehension in which they are lacking.

Previous studies have reported that Malaysian students face difficulties in shifting from a school learning culture to a university culture in meeting the academic demands at the university (Normazidah et al., 2012). Thus, information on students' abilities in certain skills such as listening would benefit both the students and instructors and this can be done by having an appropriate listening test specification that suits the students at the tertiary level.

1.7 Limitation of the Study

This study has several limitations thus generalization of the findings needs to be done cautiously. Firstly, this study focused on the first year of degree students in UiTM Negeri Sembilan. This is because the first-year students have sat for the MUET and their experiences and perceptions towards MUET listening components are important in designing and developing the listening comprehension test of the study. Besides, their perceptions of the challenges while listening to lectures where the medium of instruction is in English is important to the study.

Secondly, this study was confined to students in UiTM Negeri Sembilan. This is due to the researcher's access to the students from three branch campuses namely Kuala Pilah, Seremban, and Rembau, and the facilities such as classrooms and computer laboratories. The study required students to sit for the developed listening test in places that had proper audio systems to ensure that students could listen to the recorded listening test clearly without any disturbance.

Finally, this study covered only a public university in the country and the test was conducted only once between March to June 2018. Thus, the findings in terms of students' listening performances were limited to that particular time and are not suitable to look at trends of all tertiary students' listening performances.

1.8 Operational Definition

This study contains certain terms. The definition of the terms is briefly explained in this chapter of this thesis in order to allow the reader to make sense of what is presented in the subsequent chapters. The definitions are obtained from prior literature that can explain the terms according to how they are applied in the study. The definitions of terms used in the study are as follows:

a) Listening Test Specification (LTS)

The table of specifications is a tool used by test developers to ensure a test or assessment measures the content and thinking skills that the test is supposed to be measured (Frey, 2018). In the study, to ensure the validity of a listening comprehension test to measure students' listening proficiency, a table of specification tools is developed to act as a guideline and named Listening Test Specification (LTS).

b) Listening Comprehension

Rost (2002) describes listening comprehension as a process of relating language to concepts in someone's memory and references in the real world. For language testing purposes, the process as mentioned above is explained by bottom-up and top-down approaches which can distinguish between skillful or unskillful listeners. In this study, listening comprehension is referred to as the processing of information from decoding the acoustic input until discovering the underlying meaning of what the speaker has said.

c) Listening Abilities

In the context of the MUET listening test, listening abilities are defined as the ability of understanding spoken discourse in the higher education context, covering various types of oral text with various lengths and levels of complexity of the test content and language (Malaysian Examination Council, 2021). While according to Brown (2004), the listening ability is defined as the ability of the students to process auditory information correctly (incorrectly) through the spoken or written form. For the purpose of this study, the listening ability is defined as the ability to respond correctly to the information received from the spoken text.

d) Listening Skills

Listening skills in this study is derived from the findings of the preliminary study, Revised Bloom's Taxonomy, and MUET listening abilities. The listening skills used in developing a test specification in the study are: (1) the ability to distinguish between main idea and supporting details, (2) the ability to understand and derive meaning of word, phrases and sentences from the standard speech, (3) the

ability to understand the content to rephrase the standard speech, (4) the ability to analyse standard speech by distinguishing fact from opinion, and relevant from irrelevant information, and (5) the ability to understand the important points and detail to summarize the standard speech

e) Listening Comprehension Test (LCT)

A listening comprehension test is defined as an instrument that is being used to assess students' listening skills and involves a variety of skills from the lowest level to the higher level of the listening process (Geranpayeh & Taylor, 2011; S. Kathleen & Kitao, 1996). In the study, Listening Comprehension Test (LCT) is an instrument that is used to test L2 students' listening abilities. It is developed by applying the Socio-Cognitive Framework for validating listening tests. The development of the test is based on the newly developed Listening Test Specification (LTS)

f) Listening Performance Descriptor (LPD)

A descriptor is a reference tool that is not intended to be used as an assessment instrument though it can be a source for the development, and it can come in the form of checklist or a grid defining several categories at different levels (Council of Europe , 2021). In this study, the Listening Performance Descriptor (LPD) is a detailed description of specific ESL listening abilities which are adapted and tailored to the listening abilities of the Revised MUET listening component, and can-do statements of the Common European Framework of Reference (CEFR). The rationale for the development of LPD is due to the fact that there is a need to have a set of performance descriptors for interpreting the students' listening performance based on their test scores. The current descriptors available such as the Self-assessment grid of CEFR and the revised band descriptors of MUET listening components are too general and need to be adapted to suit the objectives of the test developed in this study. The LPD is intended to help students to profile their listening skills based on their scores. They can decide at which level they might be by looking at a checklist of more detailed descriptors in order to self-assess their level of proficiency.

g) Test-Type

Test-type is a test format that refers to the layout of the test (Geranpayeh & Taylor, 2011). It shows the type of questions that are being used in the test. In this study, test-type is referred to the type of questions used in the listening comprehension test. Examples of test type are multiple-choice questions, map labelling, matching responses, and short answers. In the study, the test types that have been selected to be included and the test specification to develop a listening

comprehension test are multiple-choice questions (MCQ), True and False, fill-in blanks and/or diagrams, short answers, dictation, and matching answers.

1.9 Summary of the Chapter

This chapter provides a discussion on the importance of having a test system which could be improved by the inclusion and provision of information about students' listening abilities. The grades or bands that instructors obtained from the current assessments are not able to give a true picture of the students' listening abilities. The current UiTM students were found to be slightly anxious when listening to English and they highlighted that they were not aware of their strengths and weakness in listening abilities. They would like to sit for a listening comprehension test that offers various test formats and tasks, not similar to the MUET, and they would like to improve their listening abilities based on the score interpretation. Students were also exposed to similar listening activities in the classroom, that is drilling by using samples of the MUET listening test. Currently, there are limited instruments to test listening comprehension that are suitable for students except for the MUET. Students need more exposure to various ways of learning and testing listening. Therefore, the study intends to fill the gap by investigating the elements of listening comprehension that can affect students' listening performance.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

Testing is a part of the assessment that is used by instructors to render information that assists them in making a lot of decisions related to the curriculum, the instruction, or to the learners. A standardized test is commonly used to measure how proficient students are in using particular language skills. Previous studies have highlighted the importance of aligning the test with the Common European Framework of Language (CEFR) (Cambridge English, 2014; Morrow, 2004; Figueras, 2012; North, 2009), while others pointed out that changes are necessary before implementing the CEFR into practice particularly when to use it as a base for other tests (Hulstijn, 2007; Fulcher, 2004; Wisniewski, 2018). There is no definite definition of test specification, nonetheless, a good test specification ensures the item content matches the objectives or criteria to be assessed. Developing a standardized listening comprehension test requires a well-designed test specification to ensure the credibility of tests to be developed based on that test specification.

Constructing a test to assess listening skills requires a theoretically grounded and empirically oriented process. This includes taking into account the external contextual factor and cognitive processing of the settings and the requirements of the academic tasks at hand. Thus, this chapter delineates the theoretical perspectives on listening comprehension, and approaches to testing listening that are required to design and develop a test specification.

2.2 Listening Comprehension

Nowadays, listening has been recognized as an essential skill in the language learning process. Although the English language is the second language in Malaysia,

students found listening comprehension difficult and previous studies showed that listening is one of the weak skills in English learning due to examination-oriented strategy where over-reliance on a model test, test-preparation materials are the common practice in the classroom and focusing on getting the right answers is important rather than teaching listening skills (English Language Standards and Quality Council, 2015; Robinson et al., 2014; Suchitra et al., 2013; Selamat & Sidhu, 2013; Nair, 2010). Assuming listening as a passive skill and can be taught while learning other skills should be denied. This is because, listening is in fact an active process and it involves various processes and relates with listeners' background knowledge and schemata (Geranpayeh & Taylor, 2011; Vandergrift & Goh, 2012; Vandergrift, 2007). Now, listening is recognized as an important skill for second language acquisition, and it deserves to be systematically developed as a skill.

Generally, listening comprehension can be described as a process of understanding and making sense of spoken text. relating what is being heard. The International Listening Association (ILA) produced a working definition in 1994 as follows:

Listening is the process of retrieving, constructing meaning from, and responding to spoken and/or nonverbal messages. It involves the ability to retain information, as well as to react emphatically and/or appreciatively to spoken and/or nonverbal messages.

(ILA listening Post, 1995, p.1 as cited in Witkin & Trochim, 1997)

The definition will act as a reference in the study as a general definition of listening for language testing purposes. Rost (2002) described listening comprehension as a process of relating language to the concept in one's memory and to references in the real world. In other words, the listening process is the result of the interplay between language use and one's experience to what surrounds him.

According to Buck (2001), listening process is a multi-faceted on the basis that listening construct can be defined through its classification into a large number of sub-components, and viewed from a number of different perspectives. Geranpayeh and Taylor (2013) defined listening comprehension as a cognitive process and presented the process in five levels. It begins with receiving acoustic clues and transform it into a group of syllables then matching them to word-level and relate the words to the co- text. Then the meaning is constructed, and a decision is made on the relevance of the new information- if appropriate, the new information will be integrated into a representation of the larger listening event. Therefore, from the definitions above, listening comprehension can be summarized as a process to extract the linguistic information and to be applied to different communicative purposes.

With the aforementioned definition of listening comprehension and its processes, it can be concluded that there are many processes involved in listening comprehension and there are still many remaining factors that may work to influence the effective use of such a complicated skill. There are several common problems that students face with listening comprehension. According to Buck (2001), the problems involved unknown vocabularies, unfamiliar topics, fast speech rate and unfamiliar accents. Teng (2002) on the other hands classified the problems to four factors, namely speaker factor, listener factor, stimulus factor and context factor. According to Gilakjani and Sabouri (2016), one of the main problems students experience during listening activities is that they have no background knowledge of what they are listening. The listening material can be about completely different culture or variety of fields in life which requires students to have certain background knowledge to understand the context and the listening text. According to Hien (2015) who did a study on ‘Difficulties and Strategies in Listening Comprehension’, 78.9%

of the respondents found accented speech causes a problem for them to comprehend listening text. Hien continued, this is because the respondents have been exposed to materials which are designed for teaching and listening purposes and not the real language or authentic listening materials. It is evident that developing listening comprehension requires extensive listening practices which involve students doing a lot of easy, comprehensible, and enjoyable listening practices (Read & Kukulskahulme, 2015). However, in the traditional approach of teaching listening, it is challenging to increase listening comprehension as the only access that students have to listening practices is solely on their teacher in the classroom. Furthermore, Azmi et al. (2014) stated listening skills is often neglected in language classroom, as much attention is given to structure, reading, writing and vocabulary. Azmi et al. (2014) continued, this is due to great number of teachers believes this skill will develop naturally and unconsciously during language learning process.

Prior studies help to put present research into perspective by giving place for discussion but also raise more questions such as variables involved in listening in order to test or assess it. Since listening comprehension involves many processes, it is necessary to consider various factors in testing students' listening comprehension skills. Type of response format, skills to be tested, pace of the speaker, vocabulary size were among the issues that were investigated in previous research. Not having appropriate amount of vocabulary can affect students' comprehension of what they read and listen. Thus, many of them were not able to comprehend what they heard from the audio due to lack of vocabulary in the English language (Wan Faizatul Azirah, et al., 2020). For example, research done by Angelina and Lay Huah (2017) found that, majority of the students were able to achieve adequate listening comprehension for social situations but appear to struggle with listening

comprehension in academic contexts due to the students' vocabulary size of the context. Besides that, listener's factor and speaker's factor also contribute to listening comprehension performance. Asriati (2017) in her studies found that, inability to acknowledge the colloquial words and slang, inability to understand the reduced word, inability to answer, the question which required a long answer, focus loss resulted from looking for the answer to the question and the speed of the speech delivery affected the students' listening comprehension performance.

The Baseline Study reported that listening is a rather neglected skill in the teaching process (Robinson et al., 2014), despite the prevalence of listening in language use. As a receptive skill, listening functions as a primary tool for gathering information for language learning, and in fact, in real-life situations, language learners tend to rely heavily on this skill (e.g., in getting information at airports, bus station or from the media). Generally, the purpose of listening comprehension test is to assess how well candidates are able to understand the speech of the speakers of the target language. Thus, testing listening comprehension skills needs numerous requirements including the skills to be tested, difficulty level of the tasks, type of response format and the text selection. For instance, type of response format selected in testing students' listening comprehension skills are very important, and it needs to be aligned with the purpose of the test. Alderson et al. (2001) stated that, the selection of the response format or test type to assess listening comprehension may affect how students interact with the auditory stimuli and how well they convey their understanding and may further predict how well they will perform in real-life listening. Brindley and Slatyer (2002) explored the task difficulty in ESL listening test and concluded that response formats influenced the level of difficulty of each item. Therefore, to achieve standardization, testers need a test specification to guide

them to construct the listening comprehension test so that the test can be considered valid and reliable to test the students' listening comprehension performance.

From the discussion above, there is no doubt that listening comprehension involves various stages. Thus, in order for the students to comprehend what they hear requires various processes and several factors need to be considered before constructing a listening comprehension test so that the standardization of the test can be achieved and the test will assess what it is supposed to assessed.

2.3 Overview on English Language Teaching in Higher Education Malaysia

From the 1960s onward, English language teaching, learning and testing saw a marked shift away from focusing on *knowledge about how the language system works* towards an emphasis on *the ability to use language*. With the publication of the latest *Malaysian Education Blueprint 2015-2025 (Higher Education)*, the prominent role of English language is seen in all Malaysian tertiary institutions. The medium of instructions for most of the universities in Malaysia is in English except for courses that are related to specific languages such as third languages. Almost all universities require students to pass English language proficiency courses before graduating.

English language at the tertiary level in Malaysia is seen to be a vehicle to develop the soft skills, an important policy known as *Development of Soft Skills for Institutions for Higher Learning* which was introduced by the Ministry of Higher Education in 2000. The latest Malaysian Education Blueprint 2015-2025 continues to emphasize the importance of the English language within the framework of the National Education Philosophy.

The importance of the English language in the National Education Philosophy is reflected when the language is a prerequisite requirement for admission to almost all programmes at the tertiary level in Malaysia including UiTM. The level of proficiency varies depending on the types of programmes the students choose to enroll in. At the tertiary level, the English language courses offered are general English language proficiency, English for Academic Purposes (EAP) and English for Occupational Purposes (EOP). At UiTM, general English language proficiency courses are generally for the students at the foundation and diploma levels. At the degree level, the language courses offered are structured towards achieving academic literacy at the higher education institution and workplace, thus the courses focus more on critical academic reading, writing and speaking skills.

As for the language teaching methodologies, higher education institutions in Malaysia have adopted Communicative Language Teaching (CLT) and Task-based Language Teaching (CBLT) (Too, 2017). The communicative language teaching paradigm aimed at teaching language as a means for communication has reflected the approaches to the assessment of listening. In Malaysia, all students who wish to enroll in degree programmes need to sit for a language proficiency test called MUET and pass a certain level or band depending on the programmes' requirement. The MUET test design is also influenced by CLT and aligned with the international standards scale, the CEFR.

2.3.1 The Malaysian University English Test (MUET)

The Malaysian University English Test is a language proficiency test and widely used for university admissions in Malaysia. It was started more than twenty years ago on December 22, 1998, when the Minister of Education announced that all

students who wish to enrol in a university in Malaysia starting the year 2001, will have to sit for the MUET. The test specifications of the MUET syllabus in 1999 were prefaced as follows:

The syllabus aims to equip students with the appropriate level of proficiency in English so as to enable them to perform effectively in their academic pursuits at tertiary level. Broadly, the syllabus seeks to bridge the gap in language needs between secondary and tertiary education by enhancing communicative competence, by providing the context for language use that is related to academic experience and by developing critical thinking skills through the competent use of language skills.

The test is set and run by the Malaysian Examination Council (MEM) and is recognized not only in Malaysia but also in Singapore. Since it is widely used for university admission in Malaysia, this test is taken by form six students, matriculation students, diploma holders and pre-university students who plan to pursue their studies in Malaysian universities. The MUET is administered three times a year. Since the MUET is being introduced as a prerequisite to university admission in 2001, the annually report showed a concerning issues regarding the listening component of MUET. Almost every year, the candidates who sit for the MUET showed poor performance in listening. Based on the descriptions given in the test specifications, the students are mostly in band 1 and 2 which indicate limited and very limited users. Table 2.1 shows the MUET listening report from 2017-2019.

Table 2. 1 MUET Listening Report from 2017-2019 (Statistics: Malaysian Examination Council, 2020)

BAND	2018			2018			2019		
	MAR	JULY	NOV	MAR	JULY	NOV	MAR	JULY	NOV
3	18.54	15.03	14.65	21.01	19.65	14.28	19.74	16.26	11.83
2	19.83	17.70	17.84	27.94	31.81	19.52	28.58	24.00	13.82
1	9.66	13.63	15.71	15.19	20.74	21.47	18.56	17.37	15.27