

**COGNITIVE VALIDITY OF MONOLOGIC
PRESENTATION AS AN ENGLISH ORAL
PROFICIENCY ASSESSMENT TASK**

SITI RAFIZAH FATIMAH BINTI OSMAN

UNIVERSITI SAINS MALAYSIA

2020

**COGNITIVE VALIDITY OF MONOLOGIC
PRESENTATION AS AN ENGLISH ORAL
PROFICIENCY ASSESSMENT TASK**

by

SITI RAFIZAH FATIMAH BINTI OSMAN

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

July 2020

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful. All praises be to Allah as it was His blessing that made the completion of this thesis possible. My heartfelt gratitude is also for the holy Prophet Muhammad (pbuh) for it was my thought of him that inspired me and uplifted my spirit to not give up but to finish writing this thesis.

The next two souls who deserve the biggest round of applause are none other than Associate Professor Dr Shaik Malik Mohamed Ismail, my current supervisor and Professor Dato' Dr Abdul Rashid Mohamed, my former supervisor who is now Deputy Vice-Chancellor at the Al-Bukhary University. They had never given up on me but kept me on my toes. May Allah bless both of them.

My deepest gratitude goes to my beloved mother. It would not have been possible for me to finish writing this thesis without the *do'a* and blessing from her. My special thanks are also for Puan Fazmawati Zakaria and Dr Norizul Azida Darus, my current and former Heads of the Language Department and the management team of UiTM Perlis for understanding my situation and allowing me to take leave whenever I needed to. To my colleagues at the Academy of Language Studies, UiTM Perlis who had helped me with the many rounds of data collection: Associate Professor Amreet Kaur, Puan Fazmawati, Pn Zahrullaili, Puan Raihanah, Pn Mahani, Pn Nazira, Pn Anis, Pn Surina, Dr Amizura Hanadi, En Adi Afzal, Pn Saeidah, Dr Latisha Asmaak, Pn Norhajawati, Pn Majdah, Pn Norliza, Pn Nor Azira, Pn Noorazalia, Ms Ong Elly, En Mohd Izwan, Cik Siti Sarina, Cik Yang Salehah, Cik Nur Fairuz Wahida, Cik Nor

Alifah and Cik Huzaifah, you are the angels from heaven, without whom, the completion of this study and thesis would not have been possible.

I also extend my thanks to the administration staff at the School of Educational Studies, En Mohd Faizal, En Mohd Jaya, Pn Salwani, Pn Farah Shafinaz, Pn Rohartina and En Zamani, and En Yazid of the Institute of Post Graduate for their help and support in the submission of this thesis.

May Allah shower all the personalities with blessing and success in all their undertakings.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iv
LIST OF TABLES	xi
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS	xviii
LIST OF APPENDICES	xix
ABSTRAK	xx
ABSTRACT	xxii
CHAPTER 1 INTRODUCTION	1
1.0 Introduction	1
1.1 Context of the Study	5
1.2 Background to the Problem.....	7
1.3 The Problem	11
1.4 Aim and Objectives of the Study	16
1.5 Research Questions	16
1.6 Rationale of the Study	17
1.7 Significance of the Study	19
1.7 Definitions of Terms	22
1.8 Limitations of the Study.....	26
1.9 Delimitation of the Study	28
1.10 Conclusion	28
CHAPTER 2 LITERATURE REVIEW	29
2.0 Introduction	29
2.1 Oral Proficiency	30
2.1.1 Conception of Oral Proficiency	31

2.1.2	Measurement of Oral Proficiency.....	34
	2.1.2(a) Accuracy	35
2.1.3	Oral Proficiency Assessment.....	38
	2.1.3(a) Monologic Task/Individual Presentation	41
	2.1.3(a)(i) The Pros of Monologic Presentation	43
2.1.4	Oral Proficiency Assessment Implemented in the Context of this Study.....	45
2.2	Theoretical Framework	51
2.2.1	Validity	51
	2.2.1(a) Cognitive Validity.....	52
	2.2.1(a)(i) Speaking Test Cognitive Validity.....	55
	2.2.1(a)(ii) Conception of Validity	56
2.2.2	Validation	58
	2.2.2(a) Kane’s Argument-based Validation Approach	62
	2.2.2(b) Bachman and Palmer’s Validation Approach	67
	2.2.2(b)(i) Bachman and Palmer’s (2010) Assessment Use Argument.....	68
2.2.3	Speech Production	69
	2.2.3(a) Levelt’s Speech Production Model	70
	2.2.3(b) Kormos’ (2006, 2011) Bilingual Speech Production Model.....	78
2.2.4	Planning and Cognitive Processing	83
	2.2.4(a) Pre-task Planning and Cognitive Processing.....	84
	2.2.4(b) On-line Planning and Cognitive Processing	87
2.3	Problems Faced by L2 Speakers with regard to Cognitive Processing	89
2.4	Previous Studies on Cognitive Validity	92
2.4.1	Methodology.....	94
	2.4.1(a) Document Review	95

2.4.1(b)	Audio Recording	95
2.4.2	Methods of Data Analysis	96
2.4.3	Sample Size	97
2.5	Conceptual Framework	98
2.5.1	Kane’s (2006) Argument-based Approach and Bachman and Palmer’s (2010) Assessment Use Argument (AUA).....	98
2.5.2	Weir’s (2005a) Socio-cognitive Framework for Speaking Test Validity	100
2.5.3	Speech Production Models	101
2.5.4	Planning and Task Performance	101
2.6	Conclusion	102
CHAPTER 3 METHODOLOGY		103
3.0	Introduction	103
3.1	Methods of Data Gathering.....	103
3.1.1	Research Design	103
3.1.1(a)	Case Study.....	103
3.1.1(b)	Qualitative Study.....	105
3.1.1(c)	Deductive and Inductive Approaches	106
3.1.1(d)	Research Design.....	106
3.1.2	Research Site	109
3.1.3	Population and Sampling.....	110
3.1.4	EC123 Oral Proficiency Assessment Instrument – Topic Discussion (TD).....	116
3.1.4(a)	Administration of the Topic Discussion (TD)	118
3.1.4(b)	EC123 Course – Reading Course.....	119
3.1.4(c)	Course Objectives	120
3.1.4(d)	Content Standards for Speaking Component	120
3.1.4(e)	EC123 Speaking Component Course Outcome	122

3.1.5	Data Collection Techniques.....	122
3.1.5(a)	Document Review	122
3.1.5(b)	Gathering of Respondents' Pre-task Planning Note Sheets	124
3.1.5(c)	Recording of Respondents' Oral Task Performance	126
3.2	Methods of Data Analysis	127
3.2.1	Document Analysis.....	127
3.2.1(a)	Steps 1 and 2 of Document Analysis	133
3.2.1(b)	Step 3 of Document Analysis – A Hybrid of Document Analysis and Thematic Analysis	134
3.2.1(c)	Intercoder Reliability	136
3.2.2	Analysis of Respondents' Pre-task Planning Notes and Oral Task Performance	140
3.2.2(a)	Reliability of Transcribers.....	141
3.2.2(b)	Thematic Analysis of Respondents' Pre-task Planning Notes and Oral Task Performance Transcripts.....	143
3.2.2(c)	Error Analysis of Respondents' Pre-task Planning Notes and Oral Task Performance Transcripts.....	144
3.2.2(d)	Functional Analysis of Respondents' Pre-task Planning Notes and Oral Task Performance Transcripts.....	146
3.2.3	Development of Coding Scheme for Thematic Analysis, Error Analysis and Functional Analysis.....	148
3.2.3(a)	Development of Preliminary Coding Scheme	149
3.2.3(b)	Pilot Study and Development of Updated Coding Scheme	153
3.2.3(c)	Development of Final Coding Scheme	159
3.2.3(d)	Intercoder Reliability	173
3.3	Conclusion	175

CHAPTER 4 FINDINGS.....	176
4.0 Introduction.....	176
4.1 The Processes to be Elicited by L2 Pre-planned Monologic Oral Proficiency Tasks.....	176
4.1.1 Step 3 of Document Analysis.....	176
4.1.1(a) Ellis’ (2005) Planning and Task Performance in a Second Language.....	177
4.1.1(b) Skehan’s (2014) Processing Perspectives on Task Performance.....	178
4.1.1(c) Levelt’s (1989) From Intention to Articulation - Document 1.....	179
4.1.1(d) Levelt’s (1992) Accessing Words in Speech Production: Stages, Processes and Representations - Document 2.....	181
4.1.1(e) Levelt’s (1995) The Ability to Speak: From Intentions to Spoken Words - Document 3.....	181
4.1.1(f) Levelt’s Producing A Spoken Language: A Blueprint of the Speaker (1999) – Document 4.....	183
4.1.1(g) Levelt, Roelof and Meyer’s (1999) A Thoery of Lexical Access in Speech Production – Document 5.....	186
4.1.1(h) Levelt’s Phonological Encoding in Speech Production (1999) – Document 6.....	188
4.1.1(i) Levelt’s (2000) Speech Production – Document 7.....	189
4.1.1(j) Kormos’ (2011) Speech Production and Second Language Acquisition – Kormos 1.....	196
4.1.1(k) Kormos’ (2011) Speech Production and the Cognition Hypothesis – Kormos 2.....	198
4.1.2 Step 4 of Document Analysis.....	202
4.1.3 Step 5 of Document Analysis.....	210
4.2 Elicitation of Cognitive Processes during Pre-task Planning and Oral Task Performance.....	217
4.2.1 Elicitation of Cognitive Processes during Pre-task Planning.....	218

4.2.1(a)	Employment of Conceptualisation during Pre-task Planning	222
4.2.1(a)(i)	Selection of Language	225
4.2.1(a)(ii)	Selection of Speech Acts	226
4.2.1(a)(iii)	Selection of Relevant Ideas	229
4.2.1(a)(iv)	Assignment of Propositional Shape.....	232
4.2.1(a)(v)	Organisation of Ideas.....	235
4.2.1(b)	Employment of Formulation (Grammatical Encoding and Morpho-phonological Encoding) during Pre-task Planning	243
4.2.1(b)(i)	Grammatical Encoding	247
4.2.1(b)(ii)	Morpho-phonological Encoding.....	253
4.2.1(c)	Employment of Self-monitoring during Pre-task Planning	260
4.2.2	Elicitation of Cognitive Processes during Oral Task Performance.....	268
4.2.2(a)	Elicitation of Cognitive Processes during the Oral Task Performance of Pre-planned Ideas	269
4.2.2(a)(i)	Employment of Formulation (Grammatical Encoding and Morpho-phonological Encoding) during the Oral Task Performance of Pre-planned Ideas	271
4.2.2(a)(ii)	Employment of Morpho-phonological Encoding during the Oral Task Performance of Pre-planned Ideas (Use of Tense Markers).....	282
4.2.2(a)(iii)	Employment of Articulation during the Oral Task Performance of Pre-planned Ideas	289
4.2.2(a)(iv)	Elicitation of Self-monitoring during the Oral Task Performance of Pre-planned Ideas	292
4.2.2(b)	Elicitation of Cognitive Processes during the Oral Task Performance of New Ideas	300

4.2.2(b)(i)	Employment of Conceptualisation during the Oral Task Performance of New Ideas	306
4.2.2(b)(ii)	Employment of Formulation during the Oral Task Performance of New Ideas	319
4.2.2(b)(iii)	Employment of Articulation during the Oral Task Performance of New Ideas.....	332
4.2.2(b)(iv)	Employment of Self-monitoring during the Oral Task Performance of New Ideas	335
4.3	Conclusion	342
CHAPTER 5 DISCUSSION AND RECOMMENDATION		343
5.0	Introduction	343
5.1	Conclusion	343
5.1.1	Elicitation of Cognitive Processes during Pre-task Planning	343
5.1.2	Elicitation of Cognitive Processes during Oral Task Performance.....	347
5.1.3	Overall Conclusion	358
5.2	Discussion	360
5.2.1	Elicitation of Cognitive Processes during Pre-task Planning.....	360
5.2.2	Elicitation of Cognitive Processes during Task Performance.....	366
5.3	Implication	373
5.4	Recommendation.....	378
5.5	Suggestions for Future Research.....	382
5.6	Conclusion	383
REFERENCES.....		384
APPENDICES		

LIST OF TABLES

		Page
Table 2.1	Saadiyah et al.'s (2007) Classifications of Linguistic Errors.....	36
Table 2.2	Monologic Presentation in High-stakes Speaking Tests.....	44
Table 2.3	Course Contents, Test Specifications and Assessment Tasks/Tasks for the Speaking Component of English Language Proficiency Courses.....	46
Table 2.4	Facets of Validity, Messick's (1989)	58
Table 2.5	Illustration of Automatic Processing in L1 Speaking	73
Table 2.6	Illustration of the Bottleneck of Time Pressure in L2 Speaking.....	92
Table 3.1	Respondents' English Proficiency Level	115
Table 3.2	EC123 Speaking Content Standards	121
Table 3.3	Steps of Reviewing Documents	124
Table 3.4	Steps of Document Analysis.....	129
Table 3.5	Document Analysis Procedure Adopted.....	131
Table 3.6	This Study's Document Analysis Procedure	132
Table 3.7	Incorporating Thematic Analysis in Document Analysis.....	135
Table 3.8	Calculation of Percent Agreement for Intercoder Reliability.....	139
Table 3.9	Calculation of Percent Agreement for Intertranscriber Reliability.....	142
Table 3.10	Steps of Error Analysis Procedure	145
Table 3.11	Operational Language Functions Checklist.....	147
Table 3.12	Sample Coding Scheme	149
Table 3.13	Preliminary Coding Scheme	150
Table 3.14	Updated Coding Scheme.....	154
Table 3.15	Final Coding Scheme	160

Table 3.16	Calculation of Percent Agreement for Intercoder Reliability.....	174
Table 4.1	Cognitive Processes Elicited during Planning	177
Table 4.2	Cognitive Processes Elicited during Pre-task Planning	178
Table 4.3	Indicators to Conceptualisation.....	179
Table 4.4	Indicators to Formulation.....	180
Table 4.5	Indicators to Self-monitoring	180
Table 4.6	Indicators to Lexical Encoding and Phonological Encoding	181
Table 4.7	Indicators to Conceptualisation.....	182
Table 4.8	Indicators to Formulation.....	182
Table 4.9	Indicators to Articulation	183
Table 4.10	Indicators to Self-monitoring	183
Table 4.11	Indicators to Conceptualisation.....	184
Table 4.12	Indicators of Grammatical Encoding, Morpho-phonological Encoding and Phonetic Encoding	185
Table 4.13	Indicators to Articulation	185
Table 4.14	Indicators to Self-perception.....	186
Table 4.15	Indicators to Lexical Selection.....	187
Table 4.16	Indicators to Articulation	187
Table 4.17	Indicators to Self-monitoring	188
Table 4.18	Indicators to Phonological Encoding and Phonetic Encoding	188
Table 4.19	Indicators to Conceptual Preparation	189
Table 4.20	Indicators to Grammatical Encoding, Morpho-phonological Encoding and Phonetic Encoding	189
Table 4.21	Indicators to Articulation	190
Table 4.22	Indicators to Self-monitoring	190
Table 4.23	Summary of Levelt's Indicators to Conceptualisation.....	191

Table 4.24	Summary of Levelt's Indicators to Formulation.....	192
Table 4.25	Summary of Levelt's Indicators to Articulation	194
Table 4.26	Summary of Levelt's Indicators to Self-monitoring	195
Table 4.27	Kormos' Indicators to Conceptualisation.....	196
Table 4.28	Kormos' Indicators to Formulation.....	197
Table 4.29	Kormos' Indicators to Self-monitoring.....	197
Table 4.30	Indicators to Kormos' Conceptualisation.....	198
Table 4.31	Indicators to Kormos' Formulation.....	199
Table 4.32	Indicators to Kormos' Self-monitoring.....	199
Table 4.33	Kormos' Indicators to Conceptualisation.....	200
Table 4.34	Kormos' Indicators to Formulation.....	201
Table 4.35	Kormos' Indicators to Self-monitoring.....	202
Table 4.36	Levelt's and Kormos' Indicators to Conceptualisation.....	203
Table 4.37	Levelt's and Kormos' Indicators to Formulation.....	206
Table 4.38	Levelt's and Kormos' Indicators to Articulation	209
Table 4.39	Merging Levelt's and Kormos' Indicators to Self- monitoring	209
Table 4.40	Finalised Indicators to Conceptualisation	211
Table 4.41	Finalised Indicators to Formulation	212
Table 4.42	Finalised Indicators to Articulation.....	213
Table 4.43	Finalised Indicators to Self-monitoring	213
Table 4.44	Matrix of Cognitive Processes	215
Table 4.45	Elicitation of Cognitive Processes during Pre-task Planning	218
Table 4.46	Respondents' Employment of Conceptualisation during Pre-task Planning by Proficiency Level.....	223
Table 4.47	Evidence of Language Selection.....	225
Table 4.48	Evidence of Speech Acts Selection.....	227

Table 4.49	Evidence of Selection of Relevant Ideas.....	229
Table 4.50	Evidence of Assignment of Propositional Shape	232
Table 4.51	Evidence of Organisation of Ideas	236
Table 4.52	Respondents' Employment of Formulation during Pre-task Planning by Proficiency Level.....	244
Table 4.53	Evidence of Lexical Encoding	248
Table 4.54	Evidence of Syntactic Encoding	251
Table 4.55	Evidence of Morpho-phonological Encoding (Use of Tense Markers).....	254
Table 4.56	Evidence of Morpho-phonological Encoding (Use of Number Markers)	256
Table 4.57	Respondents' Employment of Stage 2 Self-monitoring during Pre-task Planning by Proficiency Level	261
Table 4.58	Evidence of Monitoring Accuracy of Language Use.....	263
Table 4.59	Evidence of Monitoring Appropriacy of Ideas	265
Table 4.60	Elicitation of Cognitive Processes during the Oral Task Performance of Pre-planned Ideas	269
Table 4.61	Respondents' Employment of Formulation during the Oral Task Performance of Pre-planned Ideas by Proficiency Level	272
Table 4.62	Evidence of Lexical Encoding	275
Table 4.63	Evidence of Syntactic Encoding	280
Table 4.64	Evidence of Morpho-phonological Encoding (Use of Tense Markers).....	283
Table 4.65	Evidence of Morpho-phonological Encoding (Use of Number Markers)	285
Table 4.66	Respondents' Employment of Articulation during the Oral Task Performance of Pre-planned Ideas by Proficiency Level	290
Table 4.67	Evidence of Articulation	292
Table 4.68	Respondents' Employment of Stage 3 Self-monitoring during the Oral Task Performance of Pre-planned Ideas by Proficiency Level	293

Table 4.69	Evidence of Monitoring Accuracy of Language Use.....	295
Table 4.70	Evidence of Monitoring of Appropriacy of Ideas	298
Table 4.71	Elicitation of Cognitive Processes during the Oral Task Performance of New Ideas	301
Table 4.72	Respondents' Employment of Conceptualisation during the Oral Task Performance of New Ideas by Proficiency Level.....	306
Table 4.73	Evidence of Language Selection.....	308
Table 4.74	Evidence of Selection of Relevant Ideas.....	310
Table 4.75	Evidence of Speech Acts Selection.....	312
Table 4.76	Evidence of Propositional Shape Assignment	314
Table 4.77	Respondents' Employment of Formulation during the Oral Task Performance of New Ideas by Proficiency Level.....	319
Table 4.78	Evidence of Lexical Encoding	322
Table 4.79	Evidence of Syntactic Encoding	326
Table 4.80	Evidence of Morpho-phonological Encoding (Use of Tense Markers).....	328
Table 4.81	Evidence of Morpho-phonological Encoding (Use of Number Markers).....	329
Table 4.82	Respondents' Employment of Articulation during the Oral Task Performance of New Ideas by Proficiency Level.....	333
Table 4.83	Evidence of Articulation	334
Table 4.84	Respondents' Employment of Stage 3 Self-monitoring during the Oral Task Performance of New Ideas by Proficiency Level	336
Table 4.85	Evidence of Monitoring Accuracy of Language Use.....	337
Table 4.86	Evidence of Monitoring Appropriacy of Ideas	340
Table 5.1	Employment of Cognitive Processes during Pre-task Planning	344
Table 5.2	Similarity and Difference in the Employment of Indicators to Conceptualisation during Pre-task Planning	345

Table 5.3	Similarity and Difference in the Employment of Indicators to Formulation during Pre-task Planning	346
Table 5.4	Similarity and Difference in the Employment of Indicators to Stage 2 Self-monitoring during Pre-task Planning	346
Table 5.5	Employment of Cognitive Processes during Oral Task Performance	349
Table 5.6	Comparison in the Employment of Cognitive Processes during Pre-task Planning and Oral Task Performance	349
Table 5.7	Similarity and Difference in the Employment of Indicators to Conceptualisation during the Oral Task Performance of New Ideas	351
Table 5.8	Comparison in the Employment of Indicators to Conceptualisation during Pre-task Planning and Oral Task Performance of New Ideas	352
Table 5.9	Similarity and Difference in the Employment of Indicators to Formulation during the Oral Task Performance of Pre-planned Ideas	352
Table 5.10	Similarity and Difference in the Employment of Indicators to Formulation during the Oral Task Performance of New Ideas	353
Table 5.11	Comparison in the Employment of Indicators to Formulation during Pre-task Planning and Oral Task Performance	353
Table 5.12	Similarity and Difference in the Employment of Indicators to Articulation during the Oral Task Performance of Pre-planned Ideas and New Ideas	354
Table 5.13	Similarity and Difference in the Employment of Indicators to Self-monitoring during Oral Task Performance of Pre-planned Ideas and New Ideas	354
Table 5.14	Comparison in the Employment of Indicators to Self-monitoring during Pre-task Planning and Oral Task Performance	355
Table 5.15	Automaticity of Formulation and Articulation during Pre-task Planning and Oral Task Performance	357

LIST OF FIGURES

		Page
Figure 2.1	Socio-cognitive Framework for Speaking Test Validity, Weir's (2005a)	54
Figure 2.2	Speech Production Model, Levelt's 1989	75
Figure 2.3	Speech Production Model, Levelt's 1999	76
Figure 2.4	Bilingual Speech Production Model, Kormos' 2006	82
Figure 2.5	Conceptual Framework	99
Figure 3.1	Research Design	107
Figure 3.2	Research Matrix	109
Figure 3.3	Sample Question Paper	116

LIST OF ABBREVIATIONS

CV	Cognitive Validity
GD	Group Discussion
MP	Monologic Presentation
TD	Topic Discussion
MUET	Malaysian University English Test
PTP	Pre-task Planning
OTP	Oral Task Performance
PI	Pre-planned Ideas
NI	New Ideas
PP	<i>Pra-persediaan</i>
PTL	Pembentangan Tugas Lisan
PIP	<i>Pembentangan Idea Pra-persediaan</i>
PIB	<i>Pembentangan Idea Baru</i>

LIST OF APPENDICES

Appendix A	EC 123 Speaking Test Scoring Rubrics
Appendix B	EC123 Speaking Test Paper (November 2017)
Appendix C	Sample Document (Document Analysis)
Appendix D1	Evidence of Intercoder Reliability (Document Analysis – the Researcher’s)
Appendix D2	Evidence of Intercoder Reliability (Document Analysis – Coder 1)
Appendix D3	Evidence of Intercoder Reliability (Document Analysis – Coder 2)
Appendix E	Evidence of Intertranscriber Reliability
Appendix F	Evidence of Intercoder Reliability (Analysis of OTP Transcript)
Appendix G	Tally Sheet of Cognitive Processes Employed by Respondents during Pre- task Planning
Appendix H	Sample Pre-task Planning Notes
Appendix I	Tally Sheet of Cognitive Processes Employed by Respondents during the Oral Task Performance of Pre-planned Ideas
Appendix J	Sample Oral Task Performance Transcripts of Pre-planned Ideas
Appendix K	Tally Sheet of Cognitive Processes Employed by Respondents during the Oral Task Performance of New Ideas
Appendix L	Sample Oral Task Performance Transcripts of New Ideas

KESAHAN KOGNITIF PEMBENTANGAN MONOLOG SEBAGAI SATU TUGASAN PENILAIAN KEMAHIRAN LISAN BAHASA INGGERIS

ABSTRAK

Kajian kes ini bertujuan untuk menilai kesahan kognitif pembentangan monolog (PM) sebagai satu tugas penilaian kemahiran lisan Bahasa Inggeris pelajar semester 2 pengajian diploma yang mengambil kursus EC123 di institusi di mana pengkaji bertugas. Kajian ini adalah perlu kerana tiada sebarang validasi *a posteriori* dilaksanakan ke atas PM tiga semester selepas ia digunakan. Penilaian kesahan kognitif ini mensasarkan pengenpastian proses-proses kognitif yang digunakan oleh penutur bahasa kedua sewaktu terlibat dalam PM yang melibatkan pra-persediaan (PP) dan pengenpastian proses-proses kognitif yang digunakan oleh responden-responden kajian sewaktu PP dan pembentangan tugas lisan (PTL). Dapatan kepada objektif pertama yang diperoleh melalui analisis dokumen dan analisis tema dibentangkan dalam bentuk matriks. Matriks ini kemudiannya digunakan untuk memandu tiga metod analisis: analisis tema, analisis fungsi and analisis nahu yang dijalankan ke atas nota PP 24 responden dan PTL mereka bagi memperoleh jawapan kepada objektif kedua. Secara amnya, hasil kajian mendapati pada tiga fasa PM: PP, pembentangan idea disediakan sewaktu PP (PID) dan pembentangan idea baru (PIB), kesemua proses kognitif yang dianjurkan oleh teori penghasilan ucapan berjaya dielisisasi. Walaubagaimanapun, tahap penggunaan setiap satu proses adalah berbeza. Sewaktu PP, penggunaan tiga proses: formulasi (automatik), konseptualisasi (bukan automatik) dan tahap 2 semakan sendiri (bukan automatik) adalah separa lengkap. Sewaktu PID dan PIB, penggunaan formulasi, artikulasi dan tahap 3 semakan sendiri

adalah lengkap dan separa lengkap setiap satunya manakala penggunaan konseptualisasi sewaktu PIB hanya separa lengkap. Mengenai persamaan dan/atau perbezaan penggunaan proses di kalangan responden yang berbeza tahap kemahiran, hanya penggunaan 13 kod pra-sedia dari keseluruhan 29 kod pra-sedia untuk kesemua indikator didominasi oleh mereka yang berkemahiran aras tinggi. Ini membawa maksud ketidakmampuan sebahagian responden untuk menggunakan 16 indikator yang lain secara sempurna tidak semestinya berpunca dari kemahiran mereka yang beraras rendah tetapi dari faktor-faktor lain. Berkenaan automatisiti penggunaan formulasi dan artikulasi, automatisiti formulasi sewaktu PIB dan PID hanya dialami oleh seorang responden yang berkemahiran aras tinggi tetapi automatisiti artikulasi sewaktu PID dan PIB dialami oleh responden-responden yang bukan sahaja berkemahiran aras tinggi tetapi juga berkemahiran aras sederhana dan rendah. Penemuan baru yang disumbangkan oleh kajian kes ini ialah: automatisiti penggunaan formulasi tidak berlaku sewaktu PP, automatisiti artikulasi sewaktu PID dan PIB tidak hanya dialami oleh mereka yang berkemahiran aras tinggi tetapi juga oleh mereka di dua aras kemahiran yang lain yang lebih rendah dan penggunaan tahap 3 semakan sendiri sewaktu PID adalah automatik pada seorang responden yang berkemahiran tinggi dan sewaktu PIB kepada dua responden (seorang berkemahiran aras tinggi dan seorang berkemahiran aras rendah). Berkenaan kajian untuk dijalankan pada masa hadapan, dicadangkan proses validasi *a posteriori* untuk kesahan kognitif atau kesahan kognitif bersama dimensi kesahan yang lain dilaksanakan selepas semakan dan kemaskini dilakukan ke atas isi kandungan dokumen kursus EC123. Seterusnya, kajian yang dapat membangunkan satu alat pengesan yang boleh mengenalpasti proses-proses kognitif yang berlaku di dalam otak (*brain-tracking device*) dan kajian untuk mengenalpasti tempoh masa yang paling bersesuaian untuk PP juga wajar dijalankan.

COGNITIVE VALIDITY OF MONOLOGIC PRESENTATION AS AN ENGLISH ORAL PROFICIENCY ASSESSMENT TASK

ABSTRACT

This case study examined the cognitive validity (CV) of monologic presentation (MP), an English oral proficiency assessment task, which is used to assess the oral proficiency of semester 2 diploma students taking EC123 at the institution of higher learning where the researcher is affiliated to. The examination was necessary as no *a posteriori* validation had been performed to the MP three semesters after its implementation despite the need for validation of assessment instruments/tasks to be done continually and temporally. It sought to fulfil two objectives: identifying the cognitive processes that L2 pre-planned monologic oral proficiency tasks should elicit from L2 speakers and examining the cognitive processes that the MP managed to elicit from the respondents. The answer to the first objective which was obtained by reviewing and analysing a number of related documents using a hybrid of document analysis and thematic analysis procedures was presented in a matrix. Information in the matrix was then used to guide the thematic, functional and error analyses of the 24 respondents' pre-task planning (PTP) notes and oral task performance (OTP) data for the cognitive processes that the MP managed to elicit. Outcomes of the analyses revealed that generally, all the cognitive processes that were supposed to be elicited during PTP, OTP of pre-planned ideas (PI) and OTP of new ideas (NI) were elicited; but the extent to which each of them was employed at each phase varied. During PTP, the employment of the two unautomatic processes: conceptualisation (also during OTP of NI) and stage 2 self-monitoring and an automatic process, formulation, was partial.

During the OTP of PI and NI; however, employment of formulation, articulation and stage 3 self-monitoring was both complete and partial while the employment of conceptualisation during the OTP of NI was just partial. In terms of similarity and/or difference in the employment of the cognitive processes among the respondents of varying proficiency levels, it was revealed that the employment of the pre-set codes to 13 indicators of altogether 29 indicators: CL1, FLE1, FSE1, FM15/8 and SMIS4/SMIS6 during PTP, FLE1, FSE1, API1 and SMOSPI4/SMOSPI6 during the OTP of PI, and FLE1, FSE1, ANI1 and SMOSNI4/SMOSNI6 during the OTP of NI, was proficiency driven, i.e. led by the higher proficiency respondents. Employment of the pre-set codes to the remaining 16 indicators; nonetheless, was not proficiency driven, indicating that the inability of some respondents to employ the pre-set codes might not be due to their low proficiency but other factors. Lastly, in terms of automaticity, the automaticity of formulation was not experienced by any of the respondents during PTP but during the OTP of PI and NI, one of the three proficient respondents experienced it. Automaticity of articulation during the OTP of PI and NI; however, was experienced by not only the proficient respondents but also the respondents of the two other proficiency levels. Stage 3 self-monitoring which is an unautomatic process was found to have been automatically employed by a proficient respondent during PTP and a proficient and a basic respondents during the OTP of NI. With regard to future research, another *a posteriori* validation study should be attempted on the CV alone or the CV together with the other dimensions of validity of the MP after the content of EC123 course documents has been updated. It is also suggested that studies on the development of a tracking device which can provide data in the form of the cognitive processes occurring in the brain of speakers and the appropriate duration for PTP is conducted.

CHAPTER 1

INTRODUCTION

1.0 Introduction

This chapter discusses the fundamentals of this study. The fundamentals include the context of this study, the problem which prompted this study to be conducted, the background to the problem, the aims and objectives of this study together with the research questions, the reasons for this study to be conducted, the significant contribution of this study, the definitions of the essential terms as well as the limitations and delimitation of this study.

The subject of this study is an English oral proficiency assessment task named monologic presentation (MP), which together with group discussion (GD) forms an English oral proficiency assessment instrument named topic discussion (TD). The instrument has been used as an English oral proficiency assessment task globally and domestically. Domestically, it has been used as an oral proficiency assessment instrument of the Malaysian University English Test (MUET) and this tertiary learning institution's few English Language courses, the current one being EC123. This study; however, examined only the validity of the MP, particularly its cognitive validity (CV). Undeniably, CV has been less researched as opposed to other types of validity, i.e. content, context, scoring, predictive and the like. It began to gain attention only in 2005, after the introduction of Weir's (2005) Socio-cognitive Framework for Speaking Test Validity. Since then, according to Huang (2010), it has been drawing more attention and research.

CV and the other dimensions of validity of the TD ever since its use as an English oral proficiency assessment task of previous courses at this institution, and currently EC123 has been *a posteriori* validated once, i.e. examined for validity after

it was put to use, by Saidatul Akmar (2006). At that time, TD was an English oral proficiency assessment instrument of EEL20 course. The TD of EC123 course; however, has not been *a posteriori* validated after its use for three semesters. This study; therefore, decided to examine the validity of the MP of EC123 course, particularly its CV because CV, together with context validity, tops the Socio-cognitive Framework for Speaking Test Validity by Weir in 2005 (see page 54), a prominent framework for the construct of speaking test validity. The placement of CV and context validity at the top of the framework can be implied to mean that the validation of oral proficiency assessment tasks must start with either CV, context validity or both of them.

Among the high-stakes speaking tests which have been examined for CV were the International English Language Teaching System (IELTS) speaking test by Huang (2013) and Aptis Speaking Test by Yan et al. (2018). Meanwhile, among the writing and reading tests which had been examined for CV were the GEPT Advanced Writing Test in Taiwan by Chan et al. (2014), IELTS reading test by Bax (2013) and Aptis Reading Test by Brunfaut and McCray in 2015.

Cognitive validity can be seen in two perspectives. John Field (2011), a senior lecturer at the CRELLA research unit of the University of Bedfordshire, UK and a prominent researcher in the field of cognitive validity, defines CV in reference to the speaking tasks of Cambridge ESOL suite as “the extent to which the tasks in question succeed in eliciting from candidates a set of processes which resemble those employed in a real-world speaking event” and “how finely the relevant processes are graded across the levels of the suite in terms of the cognitive demands they impose upon the candidate” (p. 65). Huang (2010) and Weir et al. (2013) also have their definitions of CV but theirs are in alignment only to Field’s first definition. This study; therefore,

chose to examine the CV of the MP by Field's first definition (2011), the extent to which the MP was able to elicit from the respondents the cognitive processes which are similar to those employed in similar real-life non-test monologic speaking events as espoused by Levelt (1989, 1999) and Kormos (2006, 2011) in their speech production models, within the limited time that they had for pre-task planning (PTP) and oral task performance (OTP).

As mentioned in the preceding paragraph, the MP's CV was examined via the MP's elicitation of cognitive processes and elicitation was seen via the respondents' employment of the processes when attempting the MP. The link between elicitation and employment was established by Chan et al. (2014) in their validation of the GEPT writing test through their assertion "Any writing test tasks which are cognitively valid should *elicit* from test takers cognitive processes they would normally *employ* in non-test conditions" (p. 73) and Field's (2011) first definition of cognitive validity mentioned on the previous page. Since the MP is administered in two phases: time-bound PTP and time-bound OTP, the MP's elicitation of cognitive processes was examined in terms of the processes that the respondents managed to employ within the time constraints: two minutes for PTP and two minutes for OTP per test taker. This was because, according to Weir et al. (2013), "time pressure limits the extent to which a speaker can plan what he is going to say, or monitor or revise his own utterances" (p. 287) and thus, the extent to which the cognitive processes could be elicited was unknown. Other than that, according to Dörnyei and Scott (1997, as cited in Kormos, 2006), processing time pressure is a problem faced by many L2 speakers together with another three sources of problem in communication: resource deficits, perceived deficiencies in their own language output, and perceived deficiencies in decoding the interlocutor's message. And according to Baddeley (1986, 2003), processing time

pressure is one of the bottlenecks in L2 speaking as one's working memory capacity is limited. The inability of the MP to elicit any of the processes, if any, thus, could be associated with the aforementioned sources of problem.

To examine the processes that were elicited by the MP, the researcher did a thematic analysis (Braun & Clarke, 2006), functional analysis (Lazaraton & Frantz, 1997) and error analysis (Saadiyah et al., 2007) on the respondents' PTP notes and OTP data. The analyses involved coding the evidence found in the respondents' PTP notes and transcripts of their OTP, emulating the coding work done by Smith (2017) in his examination of cognitive validity of the National Assessment of Educational Progress grade 12 US history exam. Smith coded his participants' responses to the interview questions that he posed regarding the cognitive processes that should be elicited by the test. As for this study, the coding was done as part of the thematic analysis, functional analysis and error analysis to the processes that were employed by the respondents during PTP and OTP. It was done in reference to the coding scheme prepared by the researcher based on the information extracted from the documents on speech production models by Levelt (1989, 1999) and Kormos (2006, 2011), planning and task performance by Ellis (2005), and planning and second language performance by Skehan (2005).

The cognitive processes are labelled as levels of cognitive processing in Weir's (2005) Socio-cognitive Framework for Speaking Test Validity. The processes were extracted from the Speech Production Model for L1 Speakers by Levelt (1989, 1999) and Speech Production Model for L2 Speakers by Kormos (2006, 2011). The processes are the same for all speakers which means that all speakers (both L1 and L2) should employ the four processes of speech production: conceptualisation, formulation, articulation and self-monitoring when speaking in an ordinary non-test speaking event

or in an unpressured test speaking event. This was affirmed by Kormos (2011), “Speech production researchers all agree that language production has four important components: conceptualisation, formulation, articulation and self-monitoring” (p. 15).

Though the study by Saidatul Akmar (2006) had examined the validity of this task (MP) and GD in three dimensions: cognitive (previously known as theory-based), context and scoring; it was done in breadth (involving three dimensions of validity) and it involved a course which had phased out, EEL20. Though she examined the cognitive validity (theory-based validity), she did not analyse the oral and written data produced by the respondents. This study; on the contrary, analysed the respondents’ PTP notes and their OTP data to gauge the processes that they managed to employ in the time specified and not simply their reflection of what they thought they did.

Other than that, the researcher’s decision to examine only one dimension of validity, the cognitive validity, of the MP was in accordance with Weir’s (2005) postulation that researching only one aspect or dimension of validity instead of all the five outlined in his Socio-cognitive Framework for Speaking Test Validity is acceptable. Saidatul Akmar (2013) asserts that she did not research any of the dimensions in depth in 2006 because it was impossible for her to “allocate sufficient time focusing on any one component or element” in her 2006 study because the number of elements/items included in her research tasks for the three dimensions was already considerable.

1.1 Context of the Study

At this institution of higher learning, students of diploma and bachelor programmes are required to take English proficiency courses in their first three and two semesters respectively. They have to pass the first semester’s English course

before they can pursue the next. The students who are eligible for admission to diploma programmes are those who have met the minimum entry requirement, that is obtaining a credit in English and Mathematics and three other subjects while those who are eligible for bachelor programmes are those who have sat for the MUET and have completed either their diploma or matriculation studies, or are *Sijil Tinggi Pelajaran Malaysia* (STPM) holders. The medium of instruction for the teaching of the majority of the courses, excluding the third languages and the religious courses, is the English Language.

At this campus where the study was conducted, there is not much use of the English Language despite its use as the medium of instruction at the institution. Most of the campus activities are held in the L1 and the majority of the students write and communicate with their colleagues and lecturers in their L1. As such, the students' exposure to the L2 is very limited.

The number of students on this campus ranges from 7000 to more than 10,000 at certain time. They are enrolled in ten diploma programmes and 18 bachelor programmes offered by the Faculty of Agrotechnology and Plantation, Faculty of Applied Sciences, Faculty of Sports Science and Recreation, Faculty of Computer and Mathematical Sciences, Faculty of Business and Management, Faculty of Architecture, Planning and Surveying and Faculty of Accountancy. To enrol in the majority of the diploma programmes, the qualifying grade for the English Language is at least a C.

It is also a practice at this institution that curriculum review is done every five years. For the past 19 years, there have been three phases of review. Every time a review is done, the names of courses change together with the contents of the syllabus, scope of test specifications and sometimes assessment tasks. The changes are from

EEL10 (Preparatory English) to EL12 (Consolidating Language Skills) and EC121 (Integrated Language Skills: Listening), EEL20 (Mainstream English 1) to EL26 (Intermediate English) and EC15 (Integrated Language Skills: Reading) now renamed as EC123, and from EL25 (Mainstream English 2) to EL31 (English for Academic Purposes) and EC125 (Integrated Language Skills: Writing). The table which shows the changes in course contents, test specifications and assessment tasks can be found on pages 46 to 50. The assessment instrument and tasks for semester 2 diploma students, the TD; however, remains the same.

1.2 Background to the Problem

This section will discuss the factors that underlie the problem that prompted this study to be conducted, i.e. the root of the problem and the extent to which previous studies have investigated the problem, unearthing the gaps for subsequent studies to address (University of Southern California, 2019). The root of the problem was discussed in a funnel manner, beginning with the validation of language assessment tasks, validation of speaking ability or oral proficiency assessment tasks and validation of the cognitive dimension of speaking ability or oral proficiency tasks.

In terms of validation of language assessment tasks, it is common for many language practitioners to not validate or calibrate the assessment tasks that they are going to use or have used to assess language learners' ability to speak, write, listen and read. According to Davis and Elder (2005), "relatively few validation exercises of language tests have been undertaken in the past" (p. 802-803). This could be due to their misunderstanding that no validation is needed if an assessment task has already been proved valid elsewhere, their lack of awareness of the significance of determining the validity of an assessment task in the context it is used, their beliefs that the

assessment of certain skills does not carry as much weight or does not play a big role as the assessment of other skills in measuring test takers' English language proficiency and their unfamiliarity with challenging validation works. According to Crisp and Shaw (2012), good validation studies are surprisingly challenging to be carried out.

There are a few claims made of language assessments which were put to use but were not validated before use and/or after use. The one on before use is by Hyun (2010) who asserts that many performance tests, especially those which are not high stakes, are used without validation. There are also few high-stakes assessments which have been used without prior validation (*a priori* validation – before use) such as the College English Test (CET) and the most common reason given by test setters is the task has been widely used. They believe that if the task works elsewhere, it will surely work in their contexts too. With regard to no validation after use, Alderson and Buck (1993, as cited in Weir, 2005) postulate that that many UK language examinations have failed to demonstrate how an assessment meets the established assessment criteria standard (*a priori* evidence) despite how crucial it is to show that an assessment is highly valid at the outset of its implementation (*a posteriori* validation). The need for validation to be done more than once and before use and after use is affirmed by van der Walt and Steyn (2008, p. 203), “validation of any given test can therefore not be a one-off exercise, but should be a continual one”.

The use of speaking ability or oral proficiency assessment tasks without validation at all or without continual validation is also common in Malaysia. With regard to the Malaysian University English Test (MUET), a high-stakes test taken by students pursuing higher education at Malaysia's public institutions of higher learning, the few validation studies done on the MUET, namely, by Chan and Wong (2004) on

the reliability and validity of MUET speaking test, Rethinasamy and Chuah (2011) and Juliana and Abu Bakar (2013) on the predictive validity of MUET as a whole, Norlide et al. (2007) on the construct validity of the MUET, and Norlide et al. (2018) on the scoring validity of MUET writing test; were all one-off validation studies.

The use of speaking ability or oral proficiency assessment tasks without validation or continual validation could also be attributed to the uncertainty on the measurement of suitability or accuracy of an assessment task to be used to assess one's ability to speak English. This is due to the plentiful interpretations of speaking ability or oral proficiency. There have been on-going debates on the fundamental issues of speaking assessment ever since the first oral proficiency assessment was mooted by Cambridge ESOL in 1913 (Galaczi & French, 2011). The debates are on the aspects of speaking ability or oral proficiency that need to be measured. According to Fulcher (2003), a few tests divide speaking ability or oral proficiency into only pronunciation and intonation, accuracy and fluency, while a few others categorise it in terms of strategies, or analyse it using the methods of pragmatics or discourse analysis.

The other reason for the use of oral proficiency assessment tasks without validation or continual validation is the incoherence in the approach underlying the task of speaking ability or oral proficiency assessment. According to Bachman (2002), most tasks are concerned with the assessment of speaking performance, and thus they are task-centered, which, means that they are not concerned with oral proficiency, but with the performance, completion or fulfillment of tasks. An example of such assessment is The American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI). Construct-centred approach, on the other hand, is the total opposite. It supports the theoretical models of communicative competence and communicative language ability which gives priority to language components as

a measure of language ability and relegate performance to a tool meant to assess students' language ability (Kim, 2010). This type of performance speaking assessment addresses the assessment of performance of oral proficiency. This is widely accepted for most general purpose language performance assessments (Fulcher, 2003; Luoma, 2004; Messick, 1995; Skehan, 1998). The MP, no doubt, is in synchronisation with this approach.

With regard to cognitive processes, the majority of previous studies on them concentrated only on the product, i.e. the effects of engaging cognitive processes during PTP on students' performance in terms of accuracy, fluency and complexity (Abdi et al., 2012; Ahangari & Abdi, 2011; D'Ely, 2006; Guara-Tavares, 2008; Kawauchi, 2005; Mochizuki & Ortega, 2008; Qin, 2015; Sangarun, 2005; Tavakoli & Skehan, 2005; Tuan & Neomy, 2007; Xhafaj et al., 2011; Yuan & Ellis, 2003) instead of the processes that were engaged when attempting the task, especially a task like the MP which puts test takers under the pressure of making preparation (PTP) and delivering performance (OTP) in a limited amount of time.

The few studies which examined the elicitation of processes in testing context were those by Huang (2013) and Saidatul Akmar (2006) while the studies in non-testing context were by Ahmadian and Tavakoli (2014) and Guara-Tavares (2016), Ahmadian and Tavakoli's (2014) that was on the elicitation of cognitive processes during OTP examined the processes elicited for tasks which involved unpressured on-line planning and unpressured OTP and thus, there is no information from them on the elicitation of cognitive processes by tasks with time limitation such as the MP's time-bound PTP and time-bound OTP. Guara-Tavares' (2016) and Huang's (2013) examination of CV was alike but differed from this study. They examined test takers' use of strategies during PTP and OTP respectively but the strategies was not based on

the six cognitive processes espoused by Levelt (1989, 1999) and Kormos (2006, 2011). Saidatul Akmar's (2006); however, did examine the elicitation of cognitive processes by the processes themselves but it was rather shallow as the elicitation of the processes was not measured by the elicitation of all the specific indicators as spelt out by Levelt (1989, 1999) and Kormos (2006, 2011).

The above discussion delves into the factors underlying the problem that had prompted this study to be conducted, namely, the norm of using language assessment instruments or tasks without validation at all or without continual validation, the difficulty in validating speaking ability or oral proficiency assessment instruments or tasks, the lack of studies on cognitive validity pertaining to the processes elicited by the instruments or tasks as opposed to studies on the effects planning on test takers' output (product), the dynamics of the construct of oral proficiency, and the incoherence in the approach underlying oral proficiency assessment instruments or tasks. With all the information in hand, the following section will pursue the discussion of the 'problem'.

1.3 The Problem

The problem which prompted the researcher to examine the CV of the MP was worded as 'a practice which was not informed by research'. In the context of assessment instrument or task validity, a practice which is not informed by research is the development of an assessment instrument or task which is not in reference to any theory, model or framework, or the use of an assessment instrument or task without undergoing a *priori* and/or *posteriori* validation. Such a practice was encountered by Saidatul Akmar (2013) in her study when the test that she examined for validity was concluded as "was not theoretically driven or developed on a clear model/framework

for testing spoken language” (p. 90), making the test atheoretical. The need for assessment tasks to be validated is asserted by Brown and Abeywickrama (2010) who postulate that a good test is a test which is produced based on empirical research and development, instead of agreeing on particular standards or benchmarks set elsewhere with different groups of test takers and meant for achieving a different objective(s).

As for this study, the practice which was not informed by research was the use of the TD as an English oral proficiency assessment instrument of the institution’s second semester diploma students reading EC123 without being *a posteriori* validated after its use for three semesters. The *a priori* validation was assumed to have been carried out during the course development and/or assessment task selection and setting stage as proved by the availability of the test specifications and content standards, but there is no validation of its validity after use to examine the extent to which it meets the specifications established during *a priori* validation. With regard to the MP, the specifications that are related to CV gleaned from EC123 documents are the discourse functions that test takers need to perform when attempting the task and the cognitive processes that should be elicited from test takers when they take part in non-test speaking situations, namely, the academic setting (the setting specified for the MP as mentioned in the course syllabus) during the two-minute PTP and two-minute OTP.

Cognitive processes; however, are not mentioned in any of the EC123 course documents, including the test specifications, but since the validity of the cognitive dimension needs to be seen in terms of the elicitation of cognitive processes and the speech production theory espousing them, the researcher classified them as *a priori* evidences together with the discourse functions extracted from the EC123 course documents.

The absence of research is also seen in the implementation of the TD without addressing Saidatul Akmar's research finding in 2004. Saidatul Akmar revealed that the EEL20 MP lacked theory-based validity (now termed as CV) because of the "lack of clarity in terms of what test developers and administrators hypothesize of the students' language knowledge and abilities to be and how this is documented in the test" (p. 90) and lacked explicit information on linguistic knowledge required to fulfil the tasks efficiently in the course documents.

This is evident in the establishment of the scoring rubric. It is so superficial and vague that the instructors voiced their uncertainty of the aspect(s) to emphasise on during teaching; i.e. fluency, accuracy, or task fulfilment, aspects of oral proficiency to be assessed; administration of the assessment and scoring process. None of these is mentioned in the rubric (see Appendix A). They were unsatisfied with the test specifications which were very briefly written, the absence of guidelines which can guide them on the administration of the assessment, and the scoring rubric whose descriptors are very broad and vague. They, thus, resort to making own interpretations of the three assessment criteria: breadth and depth of task fulfilment, language and communicative ability to emphasise on and to assess; administration of the assessment and scoring process (personal communication, January 2014). Such uncertainty/arbitrariness is so common that it occurred not only among the MP examiners, but also IELTS examiners. Brown (2003) revealed that though examiners scored candidates according to the features within the band descriptors, they interpret the criteria differently and include a few personal criteria which were not specified in the band scales (especially the interactional aspects of performance and fluency).

Validation is commonly done in two phases: *a priori* and *a posteriori* and the latter has to be done continually. The two phases are essential as during *a priori* validation, the standard or specifications that an assessment task is supposed to meet (*a priori*) is established while during *a posteriori*, examination of the extent to which the standard is adhered to is carried out. The extent of adherence is measured through the gathering of evidences of the processes that are engaged during the administration of the instrument or task so that the decision on whether to retain it as it is, to improve it or to revamp it can be made. Both phases of validation are not only common but also compulsory as according to Cambridge English Language Assessment (2015), “validation is an on-going process which requires regular re-evaluation and confirmation that existing alignments continue to hold”.

All ELT practitioners need to understand that the validity of an assessment task in one context cannot be generalised to other contexts. Though the MP has been used as an English oral proficiency assessment task of EEL20 (EEL20 MP already examined for validity by Saidatul Akmar in 2006) and EEL26 diploma undergraduates and in the MUET, the MP of EC123 still deserved a validation process because it is of a different course with different learning outcomes. This is supported by Fulcher (1997) who asserts that validity is a local affair, as each test administration is unique (as cited in Weideman, 2006) and “the validity of a test does not lie in what the test designers claim; rather, they need to produce evidence to support such claims starting from the initial design process” (Weir, 2005, p. 15). In other words, valid validation procedures regardless of the stage are those which are based on gathering of evidences and the evidences must be gathered in the context where it is implemented.

ELT practitioners also need to realise that validation is not only essential with high-stakes assessment but also with internally-used or low-stakes assessment. The misconception that low-stakes assessment does not have to be validated was brushed off by Bachman (2007) who asserts that there has been an increased involvement with K-12 and classroom language assessment. Bachman (2007) also asserts that previously, there was no interest in school-based or classroom assessment, but the interest and volume of research and practice in this area has grown and this is evident in the research by Kane (2006), Leung (2004), and Rea-Dickins (2000, 2004).

This study; therefore, decided to do the first *a posteriori* validation to the EC123 MP to examine the CV of the MP as an English oral proficiency assessment task. The examination began with the identification of the cognitive processes to be elicited by L2 pre-planned monologic oral proficiency tasks from documents and the information obtained was then used by the researcher to guide her in her thematic, functional and error analyses of the respondents' PTP notes and data of respondents' OTP. The examination was to identify the cognitive processes which the MP had managed to elicit during the two-minute PTP and the two-minute OTP respectively, the extent to which each of the cognitive processes was employed by the respondents, the similarity/and or difference in the elicitation of the cognitive processes among the respondents of varying proficiency levels, and the respondent(s) who experienced the automaticity of two cognitive processes (formulation - grammatical encoding, morpho-phonological encoding and phonetic encoding, and articulation).

1.4 Aim and Objectives of the Study

The main aim of this study was to examine the cognitive validity of the MP as an English oral proficiency assessment task of this institution's semester 2 diploma students.

Specifically, the objectives were:

1. to identify the cognitive processes that a pre-planned monologic English oral proficiency task should elicit from speakers
2. to examine the cognitive processes that the MP managed to elicit from the respondents

1.5 Research Questions

This research had addressed the following questions:

- I. What are the cognitive processes that a PPI monologic English oral proficiency task should elicit from L2 speakers?
 - A. What are the cognitive processes that a pre-planned monologic English oral proficiency task should elicit from L2 speakers during PTP?
 - B. What are the cognitive processes that a pre-planned monologic English oral proficiency task should elicit from L2 speakers during OTP?
- II. What were the cognitive processes that the MP managed to elicit from the respondents?
 - A. What were the cognitive processes that the the respondents managed to elicit during the two-minute PTP?
 1. To what extent were the respondents able to employ each of the cognitive processes during the two-minute PTP?

2. What was the similarity and/or difference in the employment of each of the cognitive processes from the respondents of varying proficiency levels during the two-minute PTP?
 3. Which respondent(s) experienced the automaticity of formulation during the two-minute PTP?
- B. What were the cognitive processes that the MP managed to elicit from the respondents during its two-minute OTP?
1. To what extent were the respondents able to employ each of the cognitive processes during the two-minute OTP?
 2. What was the similarity and/or difference in the employment of each of the cognitive processes from the respondents of varying proficiency levels during the two-minute OTP?
 3. Which respondent(s) experienced the automaticity of formulation and articulation during the two-minute OTP?

1.6 Rationale of the Study

The examination of the CV of the MP was necessary for several reasons.

First and most importantly, the EC123 TD, inclusive of the MP, had never been *a posteriori* validated for CV and the other dimensions of validity. And thus, there was no information on the cognitive processes that the MP managed and did not manage to elicit from test takers within the time allocated for PTP and OTP after it was put to use with EC123 students about three years ago (three semesters before the data were collected and three semesters after that). The absence of such validation; therefore, made the validity of the MP after its use, particularly the CV, unknown.

Though the EC123 course is not an oral proficiency-based course but a reading-based course and the weighting of the assessment of oral proficiency is only 15% of the total score, the MP as an oral proficiency assessment task had to be validated because any form of assessment regardless of whether it is a low or high-stakes assessment and whether it is an external or internal assessment needs assurance that it assesses what is supposed to be assessed to ensure that the scores awarded to students are the true depiction of their skills or abilities. Other than that, it was the MP and not any other oral proficiency assessment tasks that was validated because it has been used in this institution for more than 20 years with semester 2 diploma students, has been used as the MUET oral proficiency assessment task since 2001 and is widely used as an assessment task of oral proficiency for both low and high-stakes assessment.

Secondly, the negative findings obtained from the *a posteriori* validation of EEL20 MP by Saidatul Akmar in 2006 in terms of scoring and theory-based (known as cognitive validity) validity were still not dealt with by the test setting team at this institution's Academy of Language Studies. Rectifications were not made to the scoring rubric and scoring procedure. It was; therefore, the hope of the researcher to share the findings of this research with the EC123 curriculum planners and test setting team so that they can be used to improve certain aspects of the task, namely, the content of the syllabus, task setting (time limitation), task demands (discourse functions) and administration (standardisation of time limitation administered among examiners) of the MP and the instructional approach .

Thirdly, the examination of the elicitation of cognitive processes by the MP involving respondents of varying proficiency levels was essential because according to Huang (2013), "The understanding and use of cognitive validity to measure and understand task demands, and actual cognitive processes, as well as the ways they

differ across various learner groups are an essential, yet often overlooked and rarely understood”.

Finally, this *a posteriori* validation of the MP’s CV was essential as it was in support of the claim made by Sandlund et al. (2016) in their review of L2 oral proficiency assessment that there is a growing interest in assessment, validation and rater perspectives. Besides, Liu and Jia (2017) made a similar claim on the need for more studies on validation as very few endeavours have been made thus far, to validate university-based speaking assessments. The validation done by this study to the MP, a university-based L2 oral proficiency assessment task, which began with the extraction of indicators to each of the cognitive processes from documents and the presentation of them in a matrix can benefit the instructors and students. This is because, if the EC123 curriculum developers agree that the indicators be assessed during the MP, the instructors can impart the knowledge about the indicators to their students. With the knowledge, the students will know what to expect from the assessment, what to do during PTP and OTP and can prepare themselves better for the assessment.

1.7 Significance of the Study

This study which aimed to examine the CV of EC123 MP as an English oral proficiency assessment task of this institution’s semester 2 diploma students was significant in several means.

Most significantly, this study’s document analysis had managed to identify, extract and present in a systematic manner (in a matrix) the cognitive processes, sub-processes and their respective indicators which a speaker should employ when attempting an L2 PPI monologic oral proficiency task and develop a coding scheme that can be referred to, to code the evidence of the employment of the cognitive

processes that are engaged while attempting other L2 oral proficiency tasks of the same format. Other than that, this study also uncovered abundance of new codes (emergent codes) for the various ways each of the cognitive processes was or was not employed and the codes were included in the coding scheme.

Besides that, the three modes of qualitative data analyses performed to the respondents' verbatim written data from PTP notes and recorded verbatim OTP data were distinct from the data analysis procedures used in previous cognitive validity studies of oral proficiency assessment tasks. According to Sandlund et al. (2016), qualitative analysis procedures were not attempted by many other empirical studies of L2 oral proficiency testing. Taylor (2011) claims that the previous *a posteriori* validation of the cognitive dimension of an oral test task involved only the use of statistical data which was rather traditional as it only involved the gathering of evidence via statistical analysis of scores and no others. Thus far, thematic analysis procedure was only used in the cognitive validation of reading and writing assessment tasks, namely, those by Brunfaut and McCray (2015) of Lancaster University which examined test-takers' cognitive processes when attempting Aptis reading tasks and Moore and Chan (2018) which examined the cognitive constructs of academic writing. Smith (2017) who examined the cognitive processes engaged by test takers when attempting a History subject test also adopted the method. None of the cognitive validity studies on oral proficiency assessment tasks has used thematic, functional and/or error analysis method except that by Ahmadian et al. (2015).

The merit of doing data analysis in the qualitative way was; it not only revealed the cognitive processes that the respondents were able and not able to employ within the time limitation but also the extent to which they managed to employ each of the processes, and the various ways of employing the processes. Other than that, the

performance of the analyses to respondents' PTP notes and OTP data, and not the performance of any advanced statistical analysis procedures had managed to provide real-life proof to what the respondents did and said at the site, instead of what they or their instructors thought they did if the data were obtained via questionnaire, concurrent verbal protocol or retrospective interview. According to Yuan and Ellis (2003), not many studies provide much information of what participants do while they are performing a task and with regard to self-monitoring, having the information on the extent to which they are monitoring the production would provide insightful information on how the cognitive processes are employed.

Thirdly, in the context of this institution, this was a second study on the validation of a direct English oral proficiency assessment task after the study by Saidatul Akmar in 2006 and the first *a posteriori* validation of EC123 MP. That one round of validation; however, was insufficient even if EEL20 is still in use. Though Saidatul Akmar validated the same task, it was for a different course and she did not examine the extent to which the MP managed to elicit the cognitive processes by specific indicators as espoused by speech production theory, the extent to which each of the cognitive processes was elicited, the extent to which the respondents were able to employ each of the cognitive processes, the similarity and/or difference in the elicitation of each of the cognitive processes by the respondents of varying proficiency levels and the respondent(s) who experienced the automaticity of formulation and articulation within the constraints of two-minute PTP and two-minute OTP.

Fourth, the findings obtained from this study would serve as an eye opener to many, with regard to the problems faced by this campus semester 2 diploma students when attempting the MP. Most of the studies conducted on cognitive processes thus far, did not investigate the process, i.e. the cognitive processes engaged by examinees

but were concerned about the product as mentioned on page 10. Even the few that did so, theirs were not in line with what the researcher intended hers to be.

1.7 Definitions of Terms

The following terms are the key terms for this study.

Cognitive validity:

CV of an oral proficiency assessment instrument or task refers to the ability of the instrument or task to elicit from test takers' the cognitive processes that are similar to those employed in non-test speaking events and the fair grading of relevant cognitive processes across the tasks in terms of the cognitive demands they impose upon candidates (Field, 2011; Huang 2010). Field (2014) also postulates that cognitive validity is to be understood as the extent to which the tasks employed and the recorded content of the tests can elicit from candidates cognitive processes that resemble those employed by a proficient listener in a real-world listening event and the processes examined for elicitation have to be seen in the light of an established theory or model related to the cognitive processing that underlies the cognitive operations in real-life language use.

With regard to this study, the CV of the MP was examined for its ability to elicit from the respondents, particularly the proficient ones, the cognitive processes that they employed in transforming their thoughts into speech during the two-minute PTP and two-minute OTP. The examination was made by analysing their verbatim written data from PTP notes and verbatim oral data from OTP in comparison to the cognitive processes employed by L2 speakers when they attempt similar tasks in a non-test academic setting (MP). The models that were adopted for examining the

elicitation of the cognitive processes were the Speech Production Model for L1 Speakers by Kormos (1999) and the Speech Production Model for L2 Speakers by Kormos (2011), a model inspired by Levelt.

Cognitive processes:

There are four cognitive processes (equivalent to four stages) espoused in Levelt's (1989, 1999) and Kormos' (2006, 2011) speech production models. The processes are engaged by speakers in translating their thoughts from ideas to speech. The four processes or stages and their respective components are: conceptualisation (conceptualiser), formulation (formulator), articulation (articulator) and self-monitoring (speech comprehension system/parser).

Validation:

According to Messick (1989, 1996), validation is a procedure for gathering evidence regarding the relevance and representativeness of the content covered from the specified construct domain. McNamara (2000); on the other hand, sees validation as the process of evaluating a test to ensure that the defensibility and fairness of the test interpretations is based on the test-takers' performances. A complete cycle of validation (*a priori* and *a posteriori*), according to Kane (1992, 2006) and Bachman and Palmer (2010), would reveal the extent to which the backing/on-site evidence meet the *a priori* evidence established during the development of the task as an oral proficiency assessment task.

For this study, the *a posteriori* cognitive validation of the MP as an English oral proficiency assessment task of this institution's semester 2 diploma students was conducted to examine the cognitive validity of the MP.

A priori validation:

This type of validation involves non-empirical work and is done by test developers or setters while developing an assessment task for the purpose of ensuring that it is valid, i.e. assessing what it is supposed to assess accurately. It can also be called theoretical validation (Messick, 1989). Its purpose is “to investigate the specification of the construct and operationalization of the test” (Weir, 2005, p. 222) which can start with library and/or Internet research.

With regard to the TD, the oral proficiency assessment instrument consisting of the MP, the *a priori* validation was considered done and this was proved by the availability of EC123 syllabus, scheme of work, content standards and test specifications.

A priori evidence:

This type of evidence is defined as the abilities that the test is supposed to measure as well as how the tasks in the test represents the abilities in 'the real world' (outside the test itself) that test users are looking for (Weir, 2005).

As for this study, the *a priori* evidence was the syllabus, scheme of work, test specifications and content standards with regard to what is assessed in the MP, the information on the cognitive processes and their indicators espoused by Levelt (1989, 1999) and Kormos (2006, 2011) in their speech production models and the cognitive processes to be elicited during during PTP and OTP as postulated by Ellis (2005) and Skehan (2014). The aforementioned information on cognitive processes was accounted for as *a priori* evidence as the evidence can include information obtained from any media of research on “how other researchers have described the constructs or how test developers have specified what they intended to test” (Weir, 2005, p.223).