

DEVELOPING THE ENTRANCE EXAMINATION FOR DIPLOMA IN ENGLISH PROGRAMME

KAMARUL ARIFFIN BIN AHMAD

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2024

DEVELOPING THE ENTRANCE EXAMINATION FOR DIPLOMA IN ENGLISH
PROGRAMME

KAMARUL ARIFFIN BIN AHMAD

THESIS PRESENTED TO QUALIFY FOR
DOCTOR OF PHILOSOPHY

FACULTY OF HUMAN DEVELOPMENT
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2024



Please tick (✓)
Project Paper
Masters by Research
Master by Mixed Mode
PhD

/

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the 10 day of 09 20 24

i. Student's Declaration:

I, KAMARUL ARIFFIN BIN AHMAD (P20192001330) FROM FACULTY OF HUMAN DEVELOPMENT (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled DEVELOPING THE ENTRANCE EXAMINATION FOR DIPLOMA IN ENGLISH PROGRAMME is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Signature of the student

ii. Supervisor's Declaration:

I DR IZAZOL BINTI IDRIS (SUPERVISOR'S NAME) hereby certifies that the work entitled DEVELOPING THE ENTRANCE EXAMINATION FOR DIPLOMA IN ENGLISH PROGRAMME (TITLE) was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a * partial/full fulfillment for the conferment of DOCTOR OF PHILOSOPHY IN EDUCATION (EDUCATIONAL MEASUREMENT) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

10 SEPTEMBER 2024

Date

IZAZOL IDRIS, PhD
Senior Lecturer
Faculty of Human Development
Sultan Idris Education University
Tg-Mahm, Perak, MALAYSIA

Signature of the Supervisor

**INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES**

**BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM**

Tajuk / Title: DEVELOPING THE ENTRANCE EXAMINATION
FOR DIPLOMA IN ENGLISH PROGRAMME

No. Matrik / *Matric's No.*: P20192001330

Saya / I: KAMARUL ARIFFIN BIN AHMAD

(Nama pelajar / *Student's Name*)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-

acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
The thesis is the property of Universiti Pendidikan Sultan Idris
2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
Tuanku Bainun Library has the right to make copies for the purpose of reference and research.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
The Library has the right to make copies of the thesis for academic exchange.
4. Sila tandakan () bagi pilihan kategori di bawah / *Please tick () for category below:-*

SULIT/CONFIDENTIAL

Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*

TERHAD/RESTRICTED

Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*

TIDAK TERHAD / OPEN ACCESS



(Tandatangan Pelajar/ *Signature*)



IZAZOL IDRIS, PhD
Senior Lecturer
Faculty of Human Development
Sultan Idris Education University
Tg. Malm, Perak, MALAYSIA

(Tandatangan Penyelia / *Signature of Supervisor*)
& (Nama & Cop Rasmi / *Name & Official Stamp*)

Tarikh: 10 SEPTEMBER 2024

Catatan: Jika Tesis/Disertasi ini **SULIT @ TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

ACKNOWLEDGEMENTS

First and foremost, I thank the Almighty that without His blessings, I would not have completed this study and reached this milestone in my life. I really hope what was found through this study would benefit the future of academic measurement and evaluation especially the one in this institution and the country.

Of course, this dissertation would not be completed without the endless guidance and supports from my supervisors who patiently listened and helped me paved my way throughout the journey. Much knowledge I had gained from both of my supervisors, and I would definitely practice the knowledge shared should I supervise students in the future. Thank you so much Dr Izazol binti Idris and Dr Zahari bin Suppian, you are both an inspiration!

I would like to also extend my gratitude to my family members, ayah, mak, akak, adik, abg Redzal, Hidayat and to my nieces and nephews who motivated me without them knowing and of course their prayers that enlightened many of my paths that seemed to be darkened with challenges and predicaments.

The deepest gratitude to my sisters from different parents, Aja and Jia who have seen and wiped my tears and sweat throughout this journey. Hey Bs! This is for us! Our Victories together, to many more victories, sisters! To Fifi, thank you for your supports, encouragement and time listening to my cries, accompanying me here and there and provided me comforts throughout this bumpy ride. Not forgetting Lopi, thank you for listening to my stories and hey we made it! Also, to Sapiq, Kak Owa, Siti and many others (the list is too long) who clapped for me and cheered for me while I am braving through my PhD journey. To those who involved directly or indirectly as the respondents and experts, to those who have helped ease my journey, may God bless all of you and pay you abundantly.

Finally, I dedicate this dissertation to Almarhum Dr Shahrir bin Jamaluddin and Almarhum Dr Nabeel who had sparked my interest and prepared me with a strong foundation for educational measurement and evaluation field especially with the use of Rasch model.

Thank you! Thank you! Thank you!



ABSTRAK

Tujuan kajian ini adalah untuk membangunkan instrumen ujian kemasukan ke program Diploma Bahasa Inggeris agar ia selari dengan kurikulum Bahasa Inggeris di Malaysia yang berlandaskan kepada *Common European Framework of Reference (CEFR)*. Memandangkan instrumen dalam pentaksiran juga perlu selari dengan CEFR, ini menjadi motivasi kepada pembangunan instrumen bagi Ujian Kemasukan Program Diploma Bahasa Inggeris yang baharu. Bukti yang menjurus kepada kekurangan dokumentasi, termasuk bukti empirikal, kajian ini membangunkan dua instrumen yang mengandungi 30 item aneka pilihan yang setiap satunya mencakupi pemahaman, tatabahasa dan kosa kata bagi membantu pemilihan pelajar untuk program diploma tersebut. Model-model pembangunan instrumen oleh Gregory dan Cohen-Swerdlik diadaptasi dalam kajian ini. Teknik persampelan *snowball* digunakan bagi mengenal pasti 11 orang pakar CEFR dan teknik persampelan bertujuan pula digunakan untuk memilih 157 orang pelajar tahun satu program diploma sebagai peserta kajian. *Content Validity Indeks (CVI = .78)* dan *Content Validity Ratio (CVR = .63)* dimanfaatkan bagi pengiraan kesahan kandungan yang melibatkan persetujuan pakar-pakar CEFR supaya semua item melepasi nilai minimum yang ditetapkan. Model pengukuran Rasch diaplikasi untuk mengenal pasti unidimensi, logit kesukaran dan kebolehan ($\theta > -2.00 - < 2.00$), *differential item functioning (DIF)* serta analisis distraktor. Hasil analisis mendapati bahawa satu daripada dua instrumen yang dibangunkan menepati nilai-nilai statistik dan ciri psikometri yang diperlukan. Tujuh item dikenalpasti perlu penambahbaikan untuk kegunaan pada masa akan datang. Kajian ini menyimpulkan bahawa pemilihan kandungan yang jitu berdasarkan kepada standard CEFR dan kaedah analisis dengan model pengukuran Rasch terbukti dapat membantu kesahan sesuatu instrumen. Kajian ini diharapkan dapat menyumbang kepada pentaksiran bahasa Inggeris di Malaysia kerana model yang diadaptasi boleh disesuaikan oleh guru sekolah. Pengiraan CVI dan CVR juga dilihat sebagai tidak mustahil dengan keberadaan pakar-pakar CEFR di setiap Pejabat Pendidikan Daerah di Malaysia. Oleh hal yang demikian, kajian ini mencadangkan satu kaedah pembangunan dan analisis instrumen yang lebih sistematik selaras dengan piawaian CEFR.





DEVELOPING THE ENTRANCE EXAMINATION FOR DIPLOMA IN ENGLISH PROGRAMME

ABSTRACT

This research aimed to develop an entrance examination for Diploma in English (DIE) programme in accordance with the English language syllabus with the use of Common European Framework of Reference (CEFR). In assessment, instruments must be pitched to the correct CEFR level which motivated the development of a new instrument. With the evidence showing lack of empirical evidence, this study developed two instruments with 30 multiple-choice items each that included reading comprehension, grammar and vocabulary based on CEFR to facilitate the selection of the candidates. An instrument development model based on Gregory and Cohen-Swerdlik's models were adapted and used. Snowball sampling technique was utilised to recruit 11 CEFR experts whereby purposive sampling technique was used to select 157 samples who were among the first-year diploma students who fulfilled the requirements. Content Validity Index (CVI = .78) and Content Validity Ratio (CVR = .63) were calculated to quantify the expert agreement on the content validity of the items that must be above the baseline. Rasch model which included unidimensionality, logits for ability and difficulty ($\theta > -2.00 - < 2.00$), Differential Item Functioning (DIF) and distractor analyses were deployed. One instrument was found to fulfil all the psychometric properties. Seven items should be revised to be used in the future. This study concluded that proper development of items with strict content selection based on the CEFR and use of psychometric properties with the Rasch model were proven to be useful for validity. This study hopes to contribute to the English language assessments in Malaysia in which it is replicable for teachers in schools. The CVI and CVR calculations could also be conducted as there are many CEFR experts located at each Education District Office in Malaysia. Thus, this study is suggesting and proposing a more systematic instrument development and analyses in-line with CEFR.



CONTENTS

	Page
DECLARATION OF ORIGINAL WORK	ii
DECLARATION OF THESIS	iii
ACKNOWLEDGEMENTS	iv
ABSTRAK	v
ABSTRACT	vi
CONTENTS	vii
LIST OF TABLES	xiv
LIST OF FIGURES	xx
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Research Background	5
1.3 Statement of Problem	11
1.4 Differences of DiEEE and Malaysian University English Test (MUET)	13
1.5 Purpose of the Research	14
1.6 Objective of Research	16
1.7 Research Questions	17

1.8	Conceptual Framework	18
1.9	Research Framework	20
1.10	Operational Definitions	20
1.10.1	Latent Trait	20
1.10.2	Language Proficiency	22
1.10.3	Instrument Development	23
1.10.4	Psychometric Properties	24
1.10.5	Diploma in English Entrance Examination	25
1.10.6	Common European Framework of Reference for Languages	25
1.10.7	Item Analysis	26
1.10.8	Fit Statistics	27
1.10.9	Separation Index	29
1.10.10	Reliability Index	30
1.10.11	Item Difficulty Index	31
1.10.12	Distractor Analysis	32
1.10.13	Differential Item Functioning	33
1.10.14	Socio Economic	34
1.10.15	First-Year Diploma Students	35
1.11	Study Limitations	36
1.12	Importance of Research	38
1.13	Summary of the Chapter	39



CHAPTER 2	LITERATURE REVIEW	40
2.1	Introduction	40
2.2	Types of Item Format	41
2.2.1	Multiple Choice Item	42
2.2.2	Gap-Filling or Completion Item	45
2.3	The Validity of Items	47
2.3.1	Content Validity	50
2.3.2	Validity in Rasch Model	54
2.3.3	Factors Affecting Validity	56
2.4	The Reliability of Items	59
2.4.1	Internal Consistency	62
2.4.2	Reliability Coefficient	64
2.4.3	Reliability in Rasch Model	66
2.4.4	Factors Affecting Reliability	68
2.5	Types of Tests	71
2.5.1	Achievement Test	72
2.6	Second Language Acquisition Theory	73
2.7	Communicative Language Approach	74
2.8	Cognitive Code Language Approach	75
2.9	Gregory Test Development Process	77
2.10	Cohen-Swerdlik Test Conceptualisation	79
2.11	Haladyna Taxonomy of Developing Multiple Choice Questions	81



2.12	Language Testing	84
2.12.1	Common European Framework of Reference (CEFR)	84
2.12.2	Proficiency Test	85
2.12.3	Testing of Grammar	86
2.12.4	Testing Reading and Vocabulary	87
2.13	Theories Underlying the Study	93
2.13.1	Classical Test Theory	93
2.13.2	Generalizability Theory	96
2.13.3	Item Response Theory	98
2.14	Rasch Model of Measurement	103
2.15	Rasch model and 1 Parameter Logistic Model	109
2.16	Differential Item Functioning	111
2.17	Test Specification	116
2.18	Item Analysis	118
2.18.1	Item Difficulty	120
2.18.2	Distractor Analysis	124
2.18.3	Item Characteristic Curve	127
2.19	Types of Response in Multiple Choice Items	129
2.20	Diploma in English Entrance Examination	131
2.21	Review of Past Studies	132
2.22	Summary of the Chapter	139

CHAPTER 3	METHODOLOGY	140
3.1	Introduction	140
3.2	Research Design	140
3.3	Quantitative Design	143
3.4	Population and Sample	149
3.4.1	Expert Panels of CEFR	150
3.4.2	Students/test takers	151
3.5	Methods of Data Collection	155
3.6	Item development Model	156
3.7	Instrument	160
3.8	Methods of Analysing Data	163
3.9	Summary of the Chapter	168
CHAPTER 4	DATA ANALYSIS	170
4.1	Introduction	170
4.2	Demographic Findings	170
4.3	The Suitability of the Vocabulary, Grammar and Reading Comprehension as an Entrance Test	174
4.3.1	Unidimensionality	175
4.3.2	Local Independence	180
4.3.3	Item Fitting	182
4.4	Content Validity Index of the DiEEE	185
4.5	Person and Item Separation Indices of DiEEE	188

4.6	Item Analysis: The difficulty logits and distractor analysis for each of the items	194
4.6.1	Vocabulary Items	195
4.6.1.1	Vocabulary Items Set A	195
4.6.1.2	Vocabulary Items Set B	206
4.6.2	Grammar Items	212
4.6.2.1	Grammar Items Set A	213
4.6.2.2	Grammar Items Set B	222
4.6.3	Reading Comprehension Items	233
4.6.3.1	Reading Comprehension Items Set A	234
4.6.3.2	Reading Comprehension Items Set B	243
4.7	Differential Item Functioning	252
4.7.1	Differential Item Functioning Gender	253
4.7.2	Differential Item Functioning – Socio Economic (Place of Residence)	257
4.7.3	Differential Item Functioning – Socio Economic (Parents Education Level)	260
4.8	Is the Instrument Appropriate to Measure Candidates' Ability?	264
4.9	Summary of the Chapter	268
CHAPTER 5	DISCUSSION, CONCLUSION AND RECOMMENDATIONS	269
5.1	Introduction	269
5.2	Discussion	269

5.2.1	Research Question 1: Is the Vocabulary, Grammar and Reading instrument appropriate to be used as an entrance examination for Diploma in English programme?	269
5.2.2	Research Question 2: Does the instrument have the construct validity based on the Rasch model unidimensional analysis?	272
5.2.3	Research Question 3: What is the Content Validity Index of the Vocabulary, Grammar and Reading instrument of the Diploma in English entrance examination?	275
5.2.4	Research Question 4.1: What are the person and item separation indices of the items in the instrument based on Rasch model?	280
5.2.5	Research Questions 4.2: What are the difficulty indices in logits of the items in the instruments based on Rasch model?	284
5.2.6	Research Question 4.3: Are there any dysfunctional distractors for the items in the instrument?	287
5.2.7	Research Question 4.4: To determine the Differential Item Functioning based on gender and socio economics status.	294
5.2.8	Research Question 4.5: Is the distribution of the item difficulty appropriate to the ability based on Wright Map?	297
5.3	Conclusion	298
5.4	Contribution of the study	302
5.5	Recommendations for Future Research	303
5.6	Summary of the Chapter	306
	REFERENCES	307

LIST OF TABLES

Table No.		Page
1.1	The Analysis of the Existing DiEEE Entrance Test Instrument	12
2.1	Cut-off Scores for the Content Validity Index	53
2.2	Cut-off Value for the Content Validity Ratio	54
2.3	Summary of the Differences between CTT and IRT	102
2.4	Outputs and Action Suggested from the Item Analyses	120
2.5	Optimal Difficulty Index for 100 Multiple Choice Items (for Norm-referenced Test)	122
3.1	The Cohen-Swerdlik's 14 Questions to Define a Test	144
3.2	Haladyna's Taxonomy of Developing Multiple Choice Questions Checklist	147
3.3	Summary of Diploma in English Reading, Vocabulary and Grammar Instrument	160
3.4	Summary of Statistical Readings and Research Questions	166
3.5	Summary of Acceptable Readings (Standards)	167
4.1	Fractions of the CEFR Experts based on Gender, Position and Experience	171
4.2	Fractions of Samples According to Gender	171
4.3	Fractions of Samples According to Place of Residence	172

4.4	Fractions of Samples According to Parent's Education Level	173
4.5	Item Polarity of the DiEEE	177
4.6	Disattenuated Correlations between the Contrast of the Clusters of DiEEE Paper	179
4.7	Fit Statistics (Mean Square) for DiEEE Instrument	183
4.8	CVR Analyses of the DiEEE Instrument	185
4.9	CVI Analyses of the DiEEE Instrument	187
4.10	Vocabulary Item 1 (VA1)	195
4.11	Vocabulary Item 2 (VA2)	196
4.12	Vocabulary Item 3 (VA3)	196
4.13	Vocabulary Item 4 (VA4)	197
4.14	Vocabulary Item 5 (VA5)	197
4.15	Vocabulary Item 6 (VA6)	198
4.16	Vocabulary Item 7 (VA7)	201
4.17	Vocabulary Item 8 (VA8)	203
4.18	Vocabulary Item 9 (VA9)	205
4.19	Vocabulary Item 10 (VA10)	205
4.20	Vocabulary Item 1 (VB1)	206
4.21	Vocabulary Item 2 (VB2)	207

4.22	Vocabulary Item 3 (VB3)	207
4.23	Vocabulary Item 4 (VB4)	208
4.24	Vocabulary Item 5 (VB5)	208
4.25	Vocabulary Item 6 (VB6)	209
4.26	Vocabulary Item 7 (VB7)	209
4.27	Vocabulary Item 8 (VB8)	210
4.28	Vocabulary Item 9 (VB9)	210
4.29	Vocabulary Item 10 (VB10)	211
4.30	Summary of Decisions for Vocabulary Items for Set A and Set B	211

4.31	Grammar Item 1 (GA1)	213
4.32	Grammar Item 2 (GA2)	215
4.33	Grammar Item 3 (GA3)	215
4.34	Grammar Item 4 (GA4)	216
4.35	Grammar Item 5 (GA5)	216
4.36	Grammar Item 6 (GA6)	217
4.37	Grammar Item 7 (GA7)	217
4.38	Grammar Item 8 (GA8)	219
4.39	Grammar Item 9 (GA9)	220

4.40	Grammar Item 10 (GA10)	220
4.41	Grammar Item 1 (GB1)	222
4.42	Grammar Item 2 (GB2)	223
4.43	Grammar Item 3 (GB3)	225
4.44	Grammar Item 4 (GB4)	226
4.45	Grammar Item 5 (GB5)	226
4.46	Grammar Item 6 (GB6)	228
4.47	Grammar Item 7 (GB7)	229
4.48	Grammar Item 8 (GB8)	229
4.49	Grammar Item 9 (GB9)	231
4.50	Grammar Item 10 (GB10)	232
4.51	Summary of Decisions for Grammar Items for Set A and Set B	232
4.52	Reading Comprehension Item 1 (RA1)	234
4.53	Reading Comprehension Item 2 (RA2)	234
4.54	Reading Comprehension Item 3 (RA3)	235
4.55	Reading Comprehension Item 4 (RA4)	235
4.56	Reading Comprehension Item 5 (RA5)	236
4.57	Reading Comprehension Item 6 (RA6)	238
4.58	Reading Comprehension Item 7 (RA7)	239

4.59	Reading Comprehension Item 8 (RA8)	240
4.60	Reading Comprehension Item 9 (RA9)	242
4.61	Reading Comprehension Item 10 (RA10)	242
4.62	Reading Comprehension Item 1 (RB1)	243
4.63	Reading Comprehension Item 2 (RB2)	244
4.64	Reading Comprehension Item 3 (RB3)	244
4.65	Reading Comprehension Item 4 (RB4)	245
4.66	Reading Comprehension Item 5 (RB5)	245
4.67	Reading Comprehension Item 6 (RB6)	248
4.68	Reading Comprehension Item 7 (RB7)	249
4.69	Reading Comprehension Item 8 (RB8)	249
4.70	Reading Comprehension Item 9 (RB9)	250
4.71	Reading Comprehension Item 10 (RB10)	250
4.72	Summary of Decisions for Reading Comprehension Items for Set A and Set B	251
4.73	Differential Item Functioning of DiEEE Instrument based on Gender for Set A	253
4.74	Differential Item Functioning of DiEEE Instrument based on Gender for Set B	256
4.75	Differential Item Functioning of DiEEE Instrument based on Place of Residence for Set A	257

- 4.76 Differential Item Functioning of DiEEE Instrument based on 258
Place of Residence for Set B
- 4.77 Differential Item Functioning of DiEEE Instrument based on 261
Parents Education Level for Set A
- 4.78 Differential Item Functioning of DiEEE Instrument based on 262
Parents Education Level for Set B
- 4.79 Summary of Differential Item Functioning of the DiEEE 263
Instrument
- 4.80 Easy-Difficult Item Ratio According to Sets (Bond & Fox, as cited 265
in Mun, n.d.)

LIST OF FIGURES

Figure No.		Page No.
1.1	The Conceptual Framework	18
1.2	The Research Framework	20
2.1	Example of Guttman Scale (Scalogram)	105
2.2	Item Characteristic Curve	128
3.1	Item Development Model. Adapted from R.J. Gregory, 2014; Cohen-Swerdlik, 2009; Kamarul Ariffin Ahmad, 2018.	157
3.2	Flow charts on how to run Rasch analyses with Winsteps software adapted from Linacre (2020).	164
4.1	Nomogram for Enequal Size of Sub-groups for DIF Study	174
4.2	Unidimensionality Test of Vocabulary, Grammar and Reading Comprehension Instrument	175
4.3	Disattenuated Correlations between the Contrast of the Clusters of DiEEE Paper	178
4.4	Standardized Residual Correlation of the DiEEE Paper	181
4.5	Person Separation Index of the DiEEE Instrument	189
4.6	Item Separation Index of the DiEEE Instrument	191
4.7	Person Separation and Reliability of DiEEE Instrument	192

4.8	Pattern of Response for Item VA6	199
4.9	Pattern of Response for Item VA7	201
4.10	Pattern of Response for Item VA8	203
4.11	Pattern of Response for item GA7	218
4.12	Pattern of Response of Item GA10	221
4.13	Pattern of Response of Item GB2	224
4.14	Pattern of Response of Item GB5	227
4.15	Pattern of Response for Item GB8	230
4.16	Pattern of Response for Item RA5	237
4.17	Pattern of Response for Item RA8	240
4.18	Pattern of Response of Item RB5	246
4.19	DIF Plot for DiEEE Instrument Set A	255
4.20	Wright Map	267



CHAPTER 1

INTRODUCTION

1.1 Introduction

Education in Malaysia has put a great emphasis on human capital to develop the nation further. The government has put a lot of allocations from annual budget for education (Nik Azis Nik Pa & Noraini Idris, 2006; & Kementerian Pendidikan Malaysia, 2023) to ensure that the nation continuously produces intellectuals as well as the skilled workers. Nik Azis Nik Pa and Noraini Idris added that one of the biggest challenges that the system is facing is to ensure the quality of the graduates thus maintaining the institution prestige in order to retain public trust towards them in producing more intellectuals and skilled workers. Consequently, institutions in Malaysia must emphasise on their curriculum as a whole starting from the very basic level of education to the highest. This should comprise every aspect of education itself including the measurement and evaluation.





The government has transformed the education in Malaysia since the independence (Malaysia was then Tanah Melayu before the merged of Tanah Melayu, Sarawak and Sabah to form Malaysia). Prior to the independence, education in Malaysia (then Tanah Melayu) was isolated accordingly to the divide-and-rule by the British (Nik Azis Nik Pa & Noraini Idris, 2006). Since then, transformations were implemented with *Akta Pelajaran 1961* to *Kurikulum Bersepadu Sekolah Rendah* and *Kurikulum Bersepadu Sekolah Menengah*. Which each of the transformations put emphasis on different aspect of life and the development of the country. The latest transformation is the Malaysia Education Blueprint (2013 - 2025) where one of the emphases is on the use and the development of the English language in Malaysian education institutions where Common European Framework of Reference came into the picture in Malaysia education settings.



Common European Framework of Reference (CEFR) has been used to reform the English language teaching and learning from preschool up to tertiary education. The aim is to put Malaysian language proficiency including the English language curriculum on a scale where it can be compared to other countries who are using the same CEFR scale thus making the products of Malaysian education are having better chance of getting employed so it “strengthens Malaysia’s position in the global economy” (p. 113, Malaysia Education Blueprint 2013 - 2025, 2013).

The framework was outlined in the English Language Education Reform in Malaysia: The Roadmap 2015 – 2025 by the English Language Standards and Quality Council. This reform is based on the policy of *Memartabatkan Bahasa Melayu, Memperkukuh Bahasa Inggeris* (MBMMBI) is design to move the English language teaching and learning ecosystem in the country to be at par with the international





standards. The reform is set to be implemented based on three waves which are 1) Wave 1 (2015 – 2016) which involved the identification and planning of the conditions that needed to be reformed, 2) Wave 2 (2017 – 2020) which involved the implementation of plan from the first wave and leveraging the outcomes, and 3) Wave 3 (2021 – 2025) which involves the review of the outcome and future plans would also be conducted to consolidate the reform.

The country is now phasing into the Wave 3 in the roadmap where the government is planning to introduce additional language into the system based on the outcomes of the previous waves (Wave 1 and Wave 2). Therefore, to have a better insights towards the success of the waves, it is a wise step to prepare and develop English language proficiency testing instruments that are at par with the CEFR at various levels of education to measure the achievements of the aims; in which the major and standardised examination such as Sijil Pelajaran Malaysia (SPM) has already been developed based on the CEFR (Zuraidah Mohd Don & Mardziah Hayati Abdullah, 2019).

When it comes to tertiary education, the changes of the curriculum for English language have started with the reformation of the English language courses as to ensure that the current syllabi in the institutions have met and mapped to the CEFR. This is to establish continuity of the Malaysia Education Blueprint in all aspects of education that include teaching, learning as well as the assessment. Not only that the courses are designed to be at par to the scale, many university programmes especially at diploma and bachelor's degree levels, which are being offered to school leavers, are experiencing the wave of changes too. To date, there is one diploma programme being offered at a public university and two other private universities which is Diploma in





English (variation Diploma in English Language Studies offered by Kolej Universiti Islam Selangor and Diploma in English Language offered by Southern University College).

The Diploma in English programme is designed to fulfil the needs of the job markets in various field. In accordance to the Malaysia Education Blueprint 2013 – 2025, the Diploma in English programme leavers are expected to take up jobs in communication, tourism, information technology and also management fields. Given a wide range of fields has made the selection of the students becoming extensive aside from the increasing demands and number of applicants through the years. The structure of the programme covers from basic English language communication skills which include reading, writing, listening and speaking, to more sophisticated used of the language in corporate and business (higher band in CEFR) where the students are trained to master the communication to serve the purpose and the needs of the fields.

In order to ensure the quality and the students enrolled had fulfilled the requirements, for the Diploma in English course that is offered in the public university, an entrance test was developed and the test is divided into three parts; reading and grammar test, a writing test and an interview which serves the purpose as speaking test. The other two Diploma in English programmes offered by the private institutions mentioned earlier do not utilise any entrance examination and the selection of students is based on their SPM results.

The focus of this study will be the reading and grammar test where it consists of 30 multiple choice questions; 10 grammar items, 10 vocabulary items and 10 reading comprehension items. Based on a document analysis, the development of the instrument





had undergone expert review in linguistics but there is no evidence that the development had followed any item development model thus there is no record of the psychometric properties as well as the CEFR mapping. This study aims to develop an instrument according to Cohen with acceptable psychometric properties to be used as the entrance for the Diploma in English programme.

This chapter will discuss and explain the background of this study, the statement of problem, the purpose of the study, the objectives of this study, the research questions and the importance of conducting the study. The limitation of the study as well as the operational definitions related to the study will be explained in the later part of this chapter.



1.2 Research Background

The entrance examination especially English tests rise many issues as the test takers come from various level of backgrounds. Major concerns were reported about the fairness of the tests pertaining to various aspects of the test itself. This includes the suitability of the content as well as the skill it intends to measure, differences between gender and ethnicity and the implications of the test to the stakeholders (Puspawati, 2014; Butler & Iino, 2021; Yao, 2023). Therefore, the English entrance test ought to be developed carefully to ensure these concerns are addressed or at least being minimised.

To begin with in 1999, pertaining to English entrance test in the country, Malaysian University English Test (MUET) was introduced and later was used as the





entrance requirement to university (Lee, 2004; Majlis Peperiksaan Malaysia, 2019). The MUET is unfortunately used for first degree enrolment whereby other levels of study, including diploma level, require a different English entrance tests as requirement. Diploma in English programme on the other hand, requires the candidates to possess a certain level of English proficiency to enable them for the programme. Therefore, an entrance examination was developed by the experts to serve this purpose with the name Diploma in English Entrance Examination (DiEEE).

As the years go by, the competition to enrol to universities in Malaysia is increasing each year due to the number of students who obtained straight A's in the Sijil Pelajaran Malaysia (SPM) increases throughout the years (Kulasegaran, 2013; Southern University College, 2021). Consequently, the number of candidates applying for the Diploma in English course is increasing too. Those applying for the course must possess the sufficient knowledge and the proficiency of the English Language so that they fulfil the demand of the programme which is to prepare the workforce who are skilful in the use of the English language.

To the date of this study was written, there are three Diploma in English programmes offered by three different institutions in which one of those is a public university. The only public university that is offering Diploma in English programme is Universiti Pendidikan Sultan Idris, in Tanjong Malim, Perak. Whereby, the other two private institutions are Kolej Universiti Islam Antarabangsa Selangor (KUIS) in Bangi, Selangor and Southern University College, in Johor Bahru, Johor. It is also worth to note that, KUIS has a different name for the programme which is Diploma in English Language Studies, meanwhile Southern University College's programme is named as Diploma in English Language (previously known as Diploma in English).





Based on the aims of the programme, regardless the different name of the diploma programme offered, all of the programmes aim to equip the students with linguistic knowledge and skills as well as to prepare them for careers which involve the use of English language such as translator, editor, journalist even to prepare them for bachelor's degree in education. As aforementioned, due to its broad career prospect, the Diploma in English programme has become high in demand (Southern University College, 2021). Therefore, a careful selection of the candidates must be done to ensure the candidates possess the required skills.

Out of the three programmes, only the one offered by Universiti Pendidikan Sultan Idris is having a special entrance test whereby the Diploma in English Language Studies (KUIS) and the Diploma in English Language (Southern University College) will only screen the applicants based on the entry requirements. When entry requirements are studied, Kolej Universiti Islam Antrabangsa (2023) only requires three credits for any subject in Sijil Pelajaran Malaysia, Southern University College (2023) requires three credits for any subject including English in Sijil Pelajaran Malaysia, and Universiti Pendidikan Sultan Idris (2023) requires a minimum A- in English and minimum E in mathematics, three credits in any subject including Bahasa Melayu as well as pass History paper in Sijil Pelajaran Malaysia, and the candidates must also pass the special entrance test.

When a test is in the picture, according to Kubiszyn and Borich (2000), a test can be used to serve various purposes including selection purposes, decision making and for placement. For selection purpose, the test must at least be able to diagnose the candidate's current existing knowledge and be able to prognose candidate's performance in the future. DiEEE, which is a proficiency test in which the test is





supposed to provide insights about candidate's existing knowledge of the English language, is used in measuring the intended skills of the candidates. Hence, DiEEE should be able to yield reliable scores throughout its components, thus assisting the faculty in the selection of their Diploma in English students.

As for the entrance test, the ability to use the English language, which is the latent trait, is to be measured in the examination. In measurement and evaluation this latent trait is assigned with numbers and a verdict of the measured latent trait will be made based on the numbers. According to Osterlind (2006), latent traits include the cognitive behaviour of a person, and he added that measurement is used in order to discover those behaviours which happen cognitively. The cognitive behaviour which is meant to be tested in the DiEEE will be the knowledge and the English language skills possessed by the candidates. The knowledge and skills cannot be observed with bare eyes and therefore DiEEE is used as the measurement tool to facilitate the selection and to be considered as one of the determining sources of entrance.

When it comes to the development of the examination instrument, it must abide to rules and models suggested by many measurement experts which include the development of items, rubric as well as expert judgement, and a psychometric method must be applied to measure the effectiveness of the instrument and to ensure the overall quality of the items. By employing the psychometric method also, test developers could plan, develop and interpret test scores in many different ways to arrive to the most accurate estimation and conclusion. The use of the test scores have already helped many test developers in various fields and the generated results had helped them to improve and to ensure that the testing system was at its best (Wood & Butterworth, 1997; Brody, 1985 as cited in Osterlind 2006).





On the other hand, the psychometric properties of the items, which could be obtained through the item analysis, should be done and the investigation of the properties should be expanded to other analyses such as Differential Item Functioning (DIF) between gender as well as socio-economic and distractor analysis. There are many ways of doing so with either the application of Classical Test Theory (CTT) or the Item Response Theory (IRT). Although the discussion concerning these two test theories will invite different opinions on which one succeeds the other, it is wise also to realise that both theories co-exist and complementing each other. CTT has its own contributions in measurement field and that makes it still relevant in many research today and IRT on the other hand contributes distinctively when it comes to determining the quality of an instrument at item level.

Developing an English testing instruments could propose certain unavoidable challenges given that there is no specific theme or topic for that purpose. This is because of that the learning, as well as testing, of language is based on the skills rather than on the content. There are a few types of test that can be used by any researchers or teachers to achieve the objective and according to Hughes (2003), Gronlund (1982) and Harrison (1983) as cited in Geethanjali Narayanan (2017), there are four types of language tests and they are (1) diagnostic test where it is used to understand the behaviour or the existing knowledge of English for the takers, (2) achievement test to measure the mastery of the language, (3) placement test which is used to classify the test takers based on their language ability or knowledge, and (4) proficiency test that is used to measure test takers' general ability to a certain requirements or for future and as prognostic use – the Diploma in English Entrance Examination is classified as a





proficiency test where it tests the candidates' abilities of the English language based on the CEFR.

As a proficiency test, the Diploma in English Entrance Examination must go through validity and reliability tests to ensure the requirements are met before it is used for selection. The developer must take into account the cost efficiency and the ease of administration when the test is designed to be taken by many; based on the review of past years' tests for DiEEE, an average of 2000 candidates would sit for the test. According to McNamara (2005), there is a lack of evidence showing test validation has been done in many language tests due to the cost and also a longer period of time to do so; where a test needs to be tested and revised then retested until the desired psychometric requirements are met. Therefore, a crucial step should be taken to begin the journey of validating the instrument.



The Diploma in English Entrance Examination which is divided into three parts comprises of reading, vocabulary and grammar test, a writing test and an interview which fall under speaking test category must be developed accordingly to ensure the overall quality of the items. The focus of this research will be the reading, vocabulary and grammar test instrument which comprises of 30 multiple choice questions: 10 questions on grammar, 10 questions on vocabulary and 10 questions for reading comprehension with one reading passage.





1.3 Statement of Problem

Many researches, dated back to previous decades to the recent ones, have proven that the development of instrument without a proper procedure can lead to many problems where the test scores could underpredict or overpredict someone's ability or it could also be biased where one subgroup of the test takers may have the advantage to perform better (Naomi, 1991; Pennock-Roman, 2002; Moshinsky & Kazin, 2005; Altamis, 2016; Fedynich, 2017; Wedman, 2017; Song, 2018; Tsaousis, Sideridis & Al-Saawi, 2018; Althewini & Alkushi, 2020). DiEEE is a teacher-made test which is supposed to anchor its content to the Common European Framework of Reference (CEFR) for English language as to not only uphold the government aspirations and the roadmap but to be at par with a widely accepted standard. Therefore, the development of the instrument must follow and be informed with the CEFR standard in which the previous one was not developed with the reference to CEFR due to the fact that the reform had not been started.

With this new addition, Thiagaran and Tan (2023), which can also be observed in Sidhu et al (2018), Noor Azli and Aini Akmar (2019), Aina Hartini Mohamah Khair and Parilah Mohd Shah (2021) and Nii and Melor Md. Yunus (2022), added that teachers are still unfamiliar with the concept of CEFR thus developing and preparing the assessment that is informed with the CEFR curriculum poses a challenge which could be a threat to its validity and reliability. Thus, the Diploma in English Entrance Examination is deemed to be carefully developed with the reference to the CEFR and to serve as the standard to be referred to by many English teachers in the country. Not only that this is needed in the education setting and to uphold the institution as the main



reference for Malaysian education paradigm, but this is also needed to fill the methodological gap on the CEFR-informed teacher made test framework which is lacking in Malaysian education setting.

In addition, as mentioned by Atalmis (2016), a proper guideline should be complied to when developing an instrument, the development of the new CEFR-informed instrument has to be done in a systematic manner. Consequently, to ensure the validity, reliability and the fairness, the items developed for a test are deemed for empirical analyses to scrutinise the quality of the item thus improving the instrument as a whole (Ali Rezigalla, 2022). Based on document analysis, there was no record of the empirical data about the psychometric properties of the current instrument which raised another concern and this opens the opportunity for this research to fill the empirical gap.

An analysis was conducted by the researcher with the existing items (non CEFR-informed) and the results are presented below with the justification to the needs of the development of a new set of items.

Table 1.1

The Analysis of the Existing DiEEE Entrance Test Instrument

Mean Raw Score	Standard Deviation	Person Separation	Person Reliability
16.3	3.6	1.06	.53

According to Linacre (2020a), person separation index of <2 with reliability <0.8 indicates that the items in the instrument are not sensitive or fail to serve the purpose of differentiating the candidates according to their ability. The analysis with

all the items in DiEEE, 303 candidate's responses had showed an alarming result where the reliability is only at .53 with the separation index of 1.06. Fisher (in Linacre 2020a) added that, a reliability of .5 with separation index of 1 show that the candidates fall under one same stratum; the instrument did not manage to categorise students into at least two categories. The analysis of DIF also showed that there are three items having the tendency of biasness and given the advantage towards one gender in getting the correct answers.

Based on the problems presented, it motivates the initiative of this study where the DiEEE reading comprehension, vocabulary and grammar paper requires a proper development procedure with certain models, appropriate taxonomy and to be empirically tested. The new set of items must undergo a stringent procedure with expert's reviews to verify the content according to CEFR, and an item analysis procedure with Rasch model. These steps are vital to achieve the requirements of a valid and reliable test. Hence, this research will develop the instrument and conduct the necessary analyses to ensure the requirements are met to facilitate a good and fair selection of the students for the programme thus filling the gaps in terms of the methodology and empirical.

1.4 Differences of DiEEE and Malaysian University English Test (MUET)

The nature of these two tests is different at the targeted proficiency. According to Majlis Peperiksaan Malaysia (2019), MUET was developed and aligned to the CEFR based on communicative language approach where the emphasis was given delivering



message and less emphasis was given to the grammar (Hui & Yunus, 2023; Jiang & Paulino, 2024). The items in the MUET examination were developed to cater from as lower as Band A2 in CEFR up to highest level which is C2 (Majlis Peperiksaan Malaysia, 2019). This is to ensure that it could capture a wide array of proficiency of the test takers.

Meanwhile the DiEEE was developed based on cognitive-code approach where the lexical and grammar aspects, or the rules, of the knowledge are given the emphasis (Chastain & Woerdehoff, 1968; Medirezen, 2014). Though the technical parts of the knowledge are tested in DiEEE, they are not tested in isolation in which the test takers are tested for their ability to use the technical knowledge to make meaningful practices. In addition, it is developed to capture only test takers with the ability ranging from Band B1 to C1+ in CEFR.

Given these differences, the use MUET to benchmark the enrolment of students to the institution would not support the aim of the programme. The future students of Diploma in English are to study the language in various levels including the linguistics aspects of the English language. The contents and the tasks of the programme are demanding which require the students to have good grasp the structural, technical and lexical of the language thus the development of the new DiEEE is called for.

1.5 Purpose of the Research

The selection of students for Diploma in English programme has to be done thoroughly due to the nature of the courses offered in the programme. Referring to the programme book, the programme requires the students to undergo a numerous number of major





courses ranging from basic English language skills to a much higher and extensive skills. Generally, the aims of those courses are to prepare the students for the future career where the Diploma in English leavers are expected to fill the demand of the workforce in various fields including journalism and translation.

Writing for the Media (course code BIK 1063) aims to train the students to write according to the characteristic and style for the print which the students will explore journalism and Translation I and Translation II (course code BIK 1223 and BIK 1233) aim to train the students to translate from Bahasa Melayu to English language, vice versa, with accurate and proper terminologies. These show that, with the programme duration of two and a half years, the students are expected to catch up with rapid lessons as well as complex knowledge of the English language to complete their study and to get the diploma. Hence, the selection of the students has to be made carefully to ensure those who are selected could survive successfully.

Therefore, this study hopes to provide useful insights of the DiEEE, in accordance to interpret candidate's language ability and in yielding significant scores to represent the candidate's true ability. To date, only expert judgments are used during the preparation of the items and after the reviewing of the previous examination papers, researcher found out that the instrument has been used for a few consecutive years and there is no evidence showing any empirical data relating to the quality of the items. Hence, this study aims to develop a more systematic documentation for the examination thus providing a more reliable and valid assessment to ensure a better process of selection.



The item analysis with Item Response Theory or commonly abbreviated as the IRT with the Rasch Model could further assist the test developers or the administrators to identify the psychometric properties of the test such as the index of difficulty. These indices are crucial to contract and deduce the quality of the item and to determine how far the items could help in selection of students for Diploma in English programme.

1.6 Objective of Research

The objectives of this research are as the following:

RO1: To identify the appropriateness of the Vocabulary, Grammar and Reading instrument to be used as an entrance examination for Diploma in English programme.

RO2: To estimate the content validity of the Diploma in English Entrance Examination instrument (for Vocabulary, Grammar and Reading) based on Content Validity Index with expert judgement.

RO3: To estimate the psychometric properties of the Diploma in English Vocabulary, Grammar and Reading instrument based on Rasch Model:

RO3.1: To estimate the person and item separation indices of the test

RO3.2: To estimate the construct validity of the instrument based on the Rasch model unidimensional analysis.

RO3.3: To estimate the difficulty indices in logits of the items

RO3.4: To determine any dysfunctional distractors in the items

RO3.5: To identify the Differential Item Functioning based on gender and socio economics status.

RO3.6: To estimate the appropriateness in terms of item difficulty distribution to ability based on Wright Map.

1.7 Research Questions

Based on the research objectives, the followings are the research questions this research aim to answer at the end of this study:

RQ1: Is the Vocabulary, Grammar and Reading instrument appropriate to be used as an entrance examination for Diploma in English programme?

RQ2: What is the Content Validity Index of the Vocabulary, Grammar and Reading instrument of the Diploma in English entrance examination based on the expert judgements?

RQ3: What are the estimation of the psychometric properties of the items in the Vocabulary, Grammar and Reading instrument of the Diploma in English entrance examination:

RQ3.1: What are the estimations of person and item separation indices of the items in the instrument based on Rasch model?

RQ3.2: Does the instrument have the construct validity based on the Rasch model unidimensional analysis?

RQ3.3: What are the estimations of difficulty indices in logits of the items in the instruments based on Rasch model?

RQ3.4: Are there any dysfunctional distractors for the items in the instrument?

RQ3.5: Are there any items showing Differential Item Functioning based on gender and socio economics in the instrument?

RQ3.6: Is the distribution of the item difficulty appropriate to the ability based on Wright Map?

1.8 Conceptual Framework

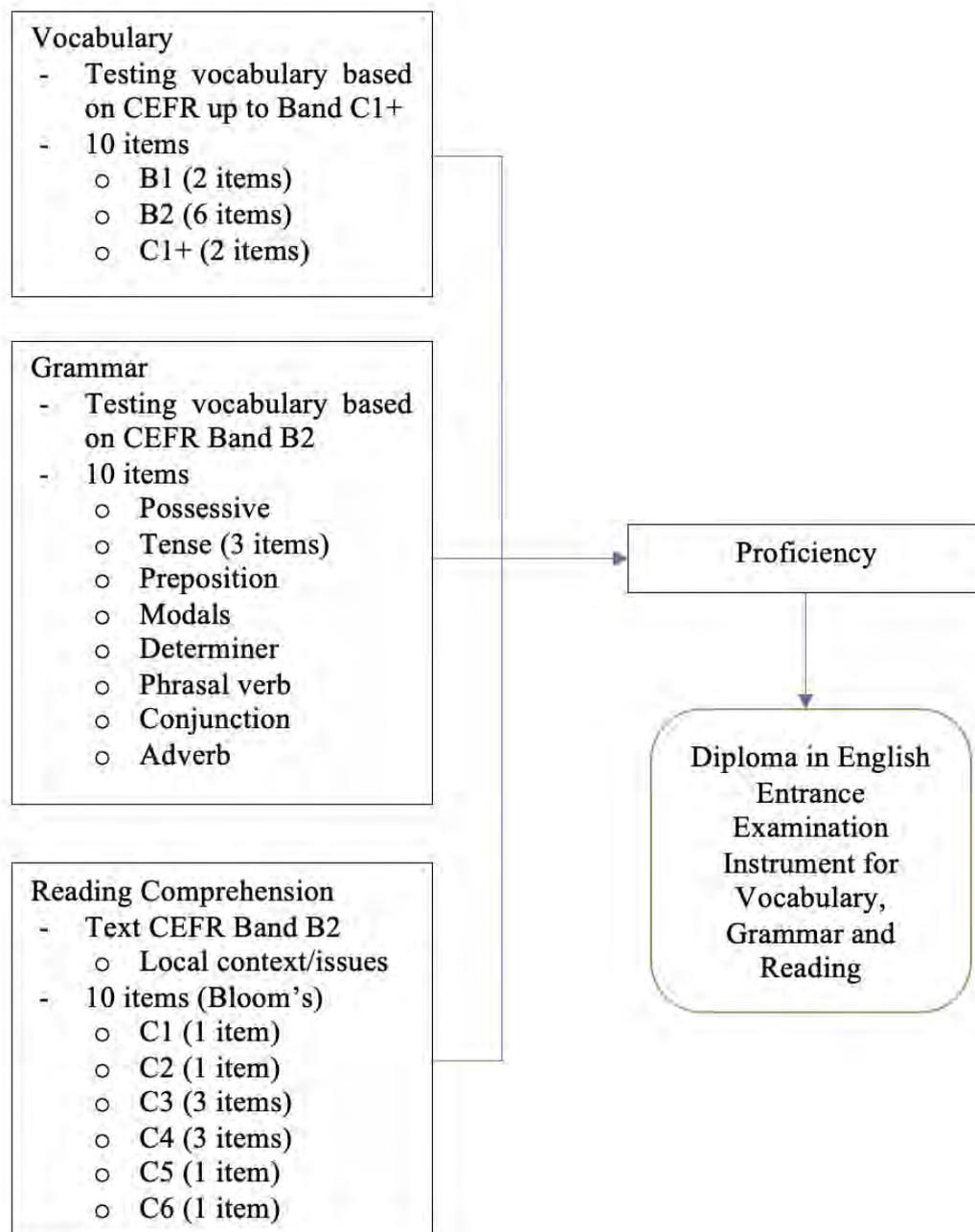


Figure 1.1. The Conceptual Framework



The Diploma in English Entrance Examination instrument comprises of three language components which are vocabulary, grammar and reading comprehension. All of the items for each of the components will be developed with Common European Framework of Reference for languages that is currently in used in the Education Blueprint 2015 – 2025. All of the items and components are designed to measure candidate's proficiency in order to facilitate the selection process of the students for Diploma in English programme.

1.9 Research Framework

The research framework shows the process the of the development and the demographic information of the research. It encapsulates the conceptualisation of the test up to the production of the instrument to be used for the Diploma in English Entrance Examination.



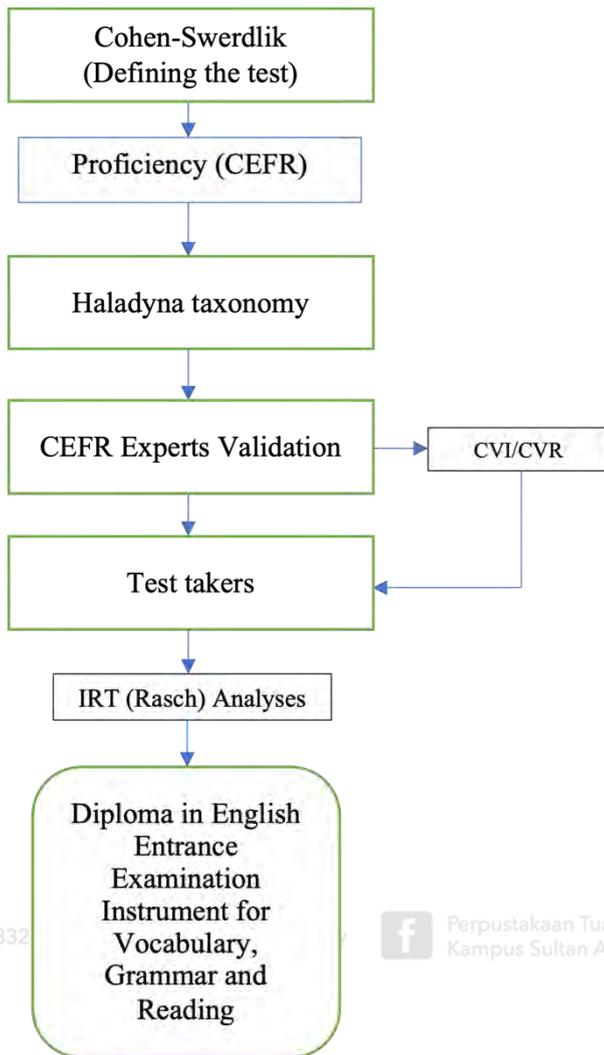


Figure 1.2. The Research Framework

1.10 Operational Definitions

1.10.1 Latent Trait

Latent traits are constructs or abilities which are unobservable with bare eyes (Hoffman, 2021) and are considered as stable but adjustable over a long period of intervention



(Measurement of Latent Traits, 2021). It may include many attributes of the respondents or the group or individual such as intelligence, attitude or aptitude. Kulkarni et al (2018) on the other hand claimed that latent traits are immovable regardless of the duration or population.

Latent traits are the constructs that distinguish people and are used to make meaningful conclusion of their personalities or abilities and since it is unobservable with bare eyes, researchers require instruments or tests to measure the traits (Fidelis, 2017) which also can be classified as proficiency of a student to complete or to function within the society (Adams, Griffin & Martin, 1987; & Fidelis, 2017). In linguistics studies, latent traits could be their knowledge of a language or their linguistic behaviours on how they use the language to deliver their meanings (Kulkarni et al, 2018) through speaking and writing.



In Rasch measurement, unidimensionality is a must to be fulfilled in order for the model to fit. This means that, Rasch model puts an emphasis on measuring only one latent trait (Bond & Fox, 2015) that can be useful to understand one's required proficiency to complete a task (Nedungadi, Paek & Brown, 2019). Therefore, this research which will be developing DiEEE instrument that comprises of grammar, vocabulary and reading comprehension sections that aim to measure one latent trait which is the proficiency; which is a constituent of their general English language knowledge.





1.10.2 Language Proficiency

Coniam and Palvey (2013) define language proficiency as the ability of a person to use the skills such as reading, listening, speaking and writing to deliver a message (Peng, Yan & Cheng, 2020). They further described that with a certain of proficiency, the language users will then be marked as competent accordingly. They also agreed that each of the skills should be tested in order for the users to be classified according to their competency levels and this is supported by Isbell and Kremmel (2020) where they added that to date many high stakes proficiency tests are deploying “a la carte approach” (p.3) especially when the tests are administered during the COVID-19 pandemic. Coombe and Davidson (2014) supported what Coniam and Palvey proposed and defined about proficiency, and they added two more dimensions to be taken into consideration when testing one’s language proficiency which are grammar and vocabulary.

Coombe and Davidson’s (2014) notion was that, without the basic knowledge and a certain size of vocabulary, one may not deliver what they intend to deliver to the audience thus making testing the vocabulary is necessary. When it comes to grammar items, they added that other than the proper and correct use of the language, patterns which is a part of grammar in a language needs to be tested too. The reason of the inclusion of these two additional dimensions in a proficiency test is based on the idea that this knowledge will help the selection purpose to be more intensive and provide insights to the faculty on which candidates would have the ability to function throughout the course of study (Coombe & Davidson, 2014; & Zhao & Liu, 2019). Shin et al. (2021) added another notion to an English proficiency test where they based the



proficiency level on the Common European Framework of Reference for English language.

In the Diploma in English Entrance Examination, the notion of the proficiency test coincides with what others have defined and described above. All of the test components in the Diploma in English Entrance Examination are aligned to what the others have said as a proficiency test that will assist the faculty in the selection of their students. Therefore, English proficiency is defined as the knowledge of the English language and the ability to make meaningful practices.

1.10.3 Instrument Development

There are several ways a researcher can adopt when he is developing an instrument. A good approach when developing an instrument must be rigorous and well structured (Horner, 2012). It may happen when a researcher aims to develop a new instrument or for modification to existing instruments and this may require a careful integration of the existing and modified version (Engineering and Physical Sciences Research Council, 2021). These approaches include phases prior to writing the items, reviewing, administering the instrument and analysing it (Davis, 1996; Meerah, et al, 2011).

Kalkbrenner (2021), claimed that there are lacking journal articles which suggest or are articulate when it comes to the topic especially the contemporary approach. Due to this, this study will adapt Gregory, Cohen-Swerdlik and Kamarul Ariffin Ahmad's frameworks in order to arrive to the conclusion of this study; where



there are steps prior to writing the item, while writing the items, analysis of pre-administration of the instrument and analysis of the responses obtain from the instrument to scrutinise its properties.

1.10.4 Psychometric Properties

Psychometric properties can be considered as the characteristics of the instrument which may describe the behaviour of the instrument. Many researchers have different opinion when it comes to the characteristics and this may include validity, reliability and responsiveness (Denman, et al, 2017), and Milne et al (2012) added another characteristic which is interpretability; though it is not considered as a core psychometric property but an important aspect to pay attention to.

Milne et al (2012) described responsiveness of an instrument as the ability of the instrument to discern changes in the response of the respondent. Whereby they described interpretability as how the data collected from the instrument can be used to make meaningful conclusion. Despite all these, De Souza, Alexandre and Guirardello (2017) claimed that reliability and validity are the two main properties that a researcher should pay attention to when using or developing an instrument. Therefore, this study will put emphasis on the validity; primarily the content validity which will be obtain from the CVI or the content validity index, and the reliability index will be obtained from the Winsteps output.





1.10.5 Diploma in English Entrance Examination

There are various English entrance examinations available throughout the world. There are International English Language Testing Systems (IELTS), Test of English as a Foreign Language (TOEFL) and Malaysian University English Test (MUET). These tests are developed by different institutions or parties that uphold and aligned to the CEFR. IELTS is developed by the British Council, TOEFL is developed by the Educational Testing System and MUET is developed by the Malaysian Examination Council. All of the mentioned tests are used as an enrolment to multiple university courses.

Diploma in English Entrance Examination is aimed to screen applicants based on specific needs of the programme thus developing a new instrument is called for. On the other hand, the Diploma in English Entrance Examination is developed based on the CEFR descriptors and to be labelled as a CEFR-informed instrument. The instrument consists of three sections with 10 multiple choice questions for each section. The sections are Vocabulary, Grammar and Reading Comprehension.

1.10.6 Common European Framework of Reference for Languages

Common European Framework of Reference for languages is an international standard that is used to benchmark the English language use globally. This framework has been adopted by many countries and they named the framework to represent the adaptations that was done to suit the context and needs within their countries (Foley, 2019). These



adaptations were named with initial of the countries to signify the locality of the framework being used for example CEFR-J (to be used in Japan) and CEFR-V (to be used in Vietnam).

The Common European Framework of Reference for Languages that is used in this study is defined and relevant to the framework that is adapted to be used in Malaysia.

1.10.7 Item Analysis

After the items are constructed, review and edited, the items will be administered to a group of test takers and the result will be analysed in a process called item analysis (Singh, 2009). Singh added that, the results obtained from the analysis will help test developers to determine whether an item will be retained, revised or even being discarded. Item analysis usually covers the readings of the item difficulty, item discrimination, distractor analysis, differential item functioning and many other statistics which deemed useful to understand the interaction between the items and the test takers. Item analysis is crucial to be deployed as many developers often spend more times developing the test than analysing the results which will at the end facilitate them to designing a better test (Wollack & Case, 2016).

Looking at the importance of performing item analyses, with Rasch by using Winsteps software, this study will deploy the analysis where the reliability of the instrument will be scrutinised as well as the separation indices, item difficulty index will be discovered, distractor analysis will be conducted, and differential item functioning will be determined.



Georg Rasch, who developed the model, believed that raw marks which are taken from a test are in ordinal scale which they are only a constituent of how many questions or items are answered correctly in the test. The marks cannot be used to indicate the test takers' true ability due to that nature and the correct answers may not be having similar difficulty indices. Rasch model of measurement calibrates the observed marks and students' ability into logits that are intervals and they can be compared directly to understand their abilities better.

As the consequence, this study will scrutinise the logits and other meaningful statistics generated from the software to answer the research questions and will not go in depth with item discrimination since Rasch model assumes the discrimination index is the same and due to the nature of the test takers who are homogenous. In this case, the fit statistics of Rasch model and the separation indices will be studied to see whether the test instrument could distinguish groups of students according to their ability.

1.10.8 Fit Statistics

Fit statistics are important to provide insights whether the items in an instrument contribute meaningfully to the construct being tested (Bond & Fox, 2007). Determining and computing the fit statistic in Rasch model can be considered as the "quality control" (Bond & Fox, p. 35) of an examination. The items with logits of between $.76$ to 1.30 are considered to be fit in the model and the items with logits outside of this range will be considered as underfit with logit 1 is considered as the perfect fit. The readings of the logits are similar in the case of person fit statistics to the model.





When the items in the instrument fit to Rasch model, the items are said to have fulfilled the assumptions of unidimensional and local independence (Brentari & Golia, 2007; Sumintono & Widhiarso, 2015). Nevertheless, Eakman (2012; cited in Shamsuddin Hasni et al, 2020) said that unidimensionality is achieved when the eigenvalue of the principal component analysis (PCA) is less than 3.0 and the percentage of the unexplained variance must be below 10% based on the output from Winsteps.

These two assumptions are important to be fulfilled before statistical procedure involving IRT, including Rasch, is to be done. When it comes to person fit statistics, those who fall outside of the said logits are considered as misfit. According to Müller (2020), those who fall $<.70$ are considered as overfit which means that the test takers belong to high achiever groups but perform badly in many low difficulty items and those who fall >1.30 are considered as underfit which indicates that the test takers belong to low achiever group but managed to answer item with high difficulty index correctly.

This study would adopt similar definition of the fit statistics to determine which item that falls outside of the acceptable logit range; with an emphasis on the item itself as this aims to develop the instrument. And the proposed technique of identifying unidimensionality by Eackmen (2012) will also be adopted in this study.





1.10.9 Separation Index

In Rasch model, particularly with Winsteps, separation indices can provide useful insights of the items and the test takers. Separation indices which are read together with the reliability, help the test developers in making decision about the overall performance of the test takers as well as the items. With Winsteps, test developers can obtain item separation index and person separation index.

According to Linacre (2020a), person separation index classifies test takers according to their ability. If the index shows 2, this indicates that the test manages to classify the test takers into two groups of ability; assumingly low achiever and high achiever. As been stated earlier, it should be read together with the reliability, Linacre further explains that an index of <2 with reliability $<.8$ suggest that the test is not sensitive enough to distinguish the two groups of test takers. This could mean that the items used in the test may all have almost similar difficulty level, or it could be due to small number of items administered in the test.

Comparable to person separation index, item separation index on the other hand classifies the items according to the level of difficulty. An index of 3 indicates that there are three level of item difficulties in the test. According to Linacre (2005, in Siti Rahayah Ariffin et al, 2010), item separation measure of >2 is considered good enough where and index of <2 is considered as not productive for measurement. On the other hand, Linacre (2020a) added that the reliability for item separation index is influenced by the sample. Reliability index of $<.9$ indicates that the number of samples is not enough for the software or for the model to locate the item according to its difficulty.



Therefore this study will follow and define good item separation index as Linacre suggested earlier where good separation index would be considered as productive for measurement.

1.10.10 Reliability Index

Producing or developing a test instrument with good, or at least acceptable, index of reliability (as well as validity) is important (Siti Rahayah Ariffin et al, 2010). Reliability can be defined as the consistency of the score of a test of an instrument across administrations (AERA, 2014). Guilford (1965; in Linacre, 1997) defined reliability as reproducibility where the instrument should yield similar results in repeated situations, and it is one of the measures of the quality of the items which the most popular measure of reliability will be Cronbach's Alpha as well as Kuder-Richardson 20 (KR-20).

According to Linacre (2020a), the equivalent value of traditional test reliability (Cronbach Alpha and KR-20) is the person reliability. Guilford (1965; in Linacre, 1997) further supported that Rasch analysis reliability is not misleading as KR-20 since KR-20 often reports a higher value or reliability (further from the true value/score). This will result that researcher or test developer believing that their instrument is reliable which in fact it is less than what is reported.

The format of the DiEEE is multiple choice questions (MCQs) and therefore the scoring of the test will be objective and dichotomous; this means that, this study will only look at the reliability of the items or the instruments and will not delve into the

inter-rater reliability. For this study, reliability means the reproducibility of the similar results throughout the candidates who will take the test and will be determined from the Winsteps output of person separation reliability. The instrument is ought to be used by the faculty for a series selection for the students of Diploma in English programme.

1.10.11 Item Difficulty Index

According to Legg (1991), difficulty index is closely related to guessing and the more difficult one item is, the greater the effect of guessing but this study puts less focus on guessing. Item difficulty index is also known as p-value (Matlock-Hatzel, 1997; Wollack & Case, 2016) which directly interprets the difficulty level of the item based on the reading of the index. The needs of conducting the analysis to find out the difficulty index is vital as a test with a lot of challenging items or a lot of easy items will yield unreliable scores (Matlock-Hatzel, 1997).

Rasch model has its own way of reporting the item difficulty index. In order to make a direct comparison between item difficulty and person's ability, Rasch model computes the observed scores to logits. Logits can range from $-\infty$ to $+\infty$ and according to Bond and Fox (2007), acceptable item difficulty are between $-2SD$ logits to $+2SD$ logits. This study defines item difficulty index as logits used in Rasch model of measurement where the negative logits would suggest easy items and positive logits indicate more difficult items.



1.10.12 Distractor Analysis

With objective items especially multiple choice items, distractors play a role in attracting low achievers to choose them. The functionality of distractors could also be used to help test developer in determining the quality of the items. Most of the time, one non-functioning distractor will be discarded but a serious attention should be given if there are too many non-functioning distractors. When the distractors fail to attract the low achievers, the difficulty index of the item will decrease thus resulting to discarding the item (Singh, 2009). Besides from that, David (2007) added that, poor distractors may result to “give away” item which could also reduce the reliability.

Rasch analysis with Winsteps has its way to help item developer to identify non-functioning distractors. According to Linacre (2020b), problematic distractor could be detected in various ways and with Winsteps, it can be detected by looking at the Outfit mean-squares of the distractors. He added that, any readings nearing 1.0 indicates that the distractors fit the model and any distractors with readings of much further than 1.0 could be considered as problematic. A distractor with an Outfit mean-square value of further than 1.0 shows that many high achievers chose that distractor thus giving assumptions that it could be the key or the correct answer too.

This study defines distractor analysis as the analysis where each of the distractors is ought to function with low achievers choosing them by studying the reading of the outfit mean-squares and the mean ability that can be found in Table 10.3 of Winsteps output.





1.10.13 Differential Item Functioning

Developing a test instrument with items which are free from any elements of biasness thus making it fair to all sub-groups of test takers is a very challenging task (Fernandes, 2015). Differential item functioning (DIF) could be identified when an item functions differently to one of the sub-groups of the test takers. The sub-group is said to interact differently to the item thus given them a higher chance in arriving to the key. In other words, DIF can be said as the advantage to a certain sub-group to answer the question correctly. Sumintono and Widhiarso (2015) define differential item functioning item as an item that is biased towards a certain group or subgroup of the test takers which possesses certain characteristics. These differences could be found between genders, across subjects or the content, socio economic status and many more.



This study will only look at two major differences which are the gender and the socio economic status. In order to identify the items with differential item functioning, the candidates of DiEEE will be given a code to represent their gender and socio economic status, and based on that a table will be generated with Winsteps to identify the items. According to Sumintono and Widhiarso (2015), items with differential item functioning will appear to have the probability index of $.05$ or less and according to Aryadoust (2017), items with differential item functioning will have a p -value of $<.05$ and the size of differential item functioning plays a role in determining one.





1.10.14 Socio Economic

Socio economic status can comprise and be categorised according to a few common measures. This may include level of education, the occupation and household income (Baker, 2014) as well as the demography – the place a family is living such as urban and rural areas (Michael, et al, 2021). Willms and Tramonte (2019) define socio economic status as the person or a family's position in a social stratum based on the aforementioned measures. They claimed that many social science researchers termed socio economic status as the occupation of the parents or an individual which is closely related to the household income. This is consistent with Skrobarcek et al (2021), where they further categorised the income into few other categories including low income household and poverty.



Household income in Malaysia can be categorised into three groups namely T20 or the Top 20% of the population, M40 or the Middle 40% of the population and B40 or the below 40% of the population. According to Shahrman Haron (Department of Statistics Malaysia, 2020), T20 being the highest rank in the socio-economic status is a group of people with a household income of more than RM9,619, followed by M40 with the household income of between RM4360 to RM9619 and B40 with the household income of lower than RM4360. Some studies suggested a different definition to the household income which they categorised the socio-economic status to two groups which are those who fall below and above poverty line. A study by Nor Azrul Mohd Zin and Noordeyana Tambi (2018) added that the poverty line could be marked at the RM3,000 especially to those who lived in the urban areas. They are said to





struggle with their survival to live and meet the basic needs of life (Nor Azrul Mohd Zin & Noordeyana Tambi, 2018; Nora'azian Nahar, 2018; & Michael, et al, 2021).

These measures of socio economy play a major role when it comes to social sciences research as certain groups in the social strata, especially the low income or the poverty group, may not receive enough or proper materials due to their penury (Webb et al, 2016). This study will look into this matter with differential item functioning following what is suggested by Michael et al (2021) which is the level of education of the parents where it will be categorized into two groups which are university graduates and non-university graduates, and the geographical demography of the family which is also categorized into two groups that are urban and rural areas based on the categories by the Kementerian Pembangunan Kerajaan Tempatan (Ministry of Local Governement Development).



1.10.15 First-Year Diploma Students

The first-year diploma students were the students who were in their first year of studies in the university and were enrolled to any of the diploma programmes offered by the university. They had sat for Sijil Pelajaran Malaysia at the end of their school year and had been selected for various diploma programmes in the varsity. Given that these programmes require different levels or English grade in the Sijil Pelajaran Malaysia, only those who obtained a minimum of grade A- will be selected in the study. This study will be using this group of students as the sample since they are already selected to enrol in the university, and they have had little to no English language lessons or



interventions in between the period of leaving schools and their enrolment thus making them a suitable group to partake in the study. Furthermore, they have already sat for the Sijil Pelajaran Malaysia thus increasing the accuracy of the selection of the sample; they had the result based on a standardized test.

1.11 Study Limitations

The limitations of the study were focused on the sampling of the test takers. In terms of sampling, the limitations arose when getting a group of samples possessing similar traits outside of an institution perimeter is merely impossible. This is due to the requirements of the programme where only high school leavers who obtain a minimum of grade A- for English language, passed Mathematics with minimum grade E, passed History subject and obtained three credits including Bahasa Malaysia in Sijil Pelajaran Malaysia (SPM) would be called to sit for DiEEE. Getting school leavers without any intervention between the SPM and the DiEEE is high unlikely. What's more to gather those who got a minimum of A- for English and fulfil all the general requirements set by the university from throughout Malaysia.

The initial plan was to conduct the test online by inviting school leavers who suit the criteria throughout Malaysia. On the account of conducting the test online, reaching the samples online may violate the results of this study since DiEEE is categorised as proficiency test and ensuring academic integrity is hard as they might receive external assistance while responding to the items. Furthermore, the ambience



could not be standardized for all the test takers thus this could seriously hamper the overall results of the study.

The instrument was then administered to fifth formers of selected MRSM (IGCSE) and their selection was based on their English IGCSE grade – IGCSE has already been aligned to CEFR thus they were seen as the potential respondents of the study as well as the potential students of the Diploma in English programme. Unfortunately, the data collected from these MRSMs showed peculiarities in the pattern of answers thus rose the questions of students' motivation when answering and responding to the instrument.

The total number of first-year diploma students who obtained a minimum of grade A- for English paper in the Sijil Pelajaran Malaysia was small. This is due to the fact that only Diploma in English programme sets a minimum of grade A- for English as the requirement whereby other diploma programmes only require a minimum of grade E or the passing grade to be selected as a student in the varsity. This strengthens the use of Rasch model of measurement, which is a robust method to perform analyses with smaller number of samples. Therefore, this study involved 157 first-year diploma students in the university as the sample of this study which comprised of students from various diploma programmes offered in the university.





1.12 Importance of Research

This study aims to provide insights and to develop the reading comprehension, vocabulary and grammar instrument for Diploma in English Entrance Examination. The development of this new instrument will give impacts to many parties at many levels of administration. Asides from that, it is also hoped that the results of this study will lead to a more substantive measurement especially when it involves decision making. Above all, the results of this study would benefit the item developers, faculty administration and for the university.

Given the fact that as long as the programme is offered by the faculty, the instrument needs to be revised once in a while and it is a wise move to develop an item bank for this purpose. Since the commencement of the programme, there is no record that the entrance examination for Diploma in English has a systematic item bank. To ensure the reliability and the validity of the instrument, the results on this study will help the item developers to tailor the items to meet the current standard. This study could also be a start for the item developers to keep the items according to the psychometric properties.

As consequence to the proper instrument development, the administration at the faculty level will also receive an impact. The selection of the Diploma in English programme students will be more accurate based on the empirical data provided by the tests. This will reduce the selection errors thus helping the faculty to select those who are really fit to enrol to the programme. Asides from assisting in the selection, the faculty will also have a proper assessment system when it comes to the selection of





students. With the development of reading comprehension, vocabulary and grammar instrument, it is hoped that the other components of the DiEEE will follow suit.

The university on the other hand, will be able to ensure that the institution learning outcome could be fulfilled. During the selection, students will be selected based on the requirements outlined by the administration and those who are selected must possess a certain level of knowledge to reduce problems throughout the study years. These are all done as a step to ensure the quality of the students during the programme and upon leaving the university. With the development of this instrument, the selection process will be incisive. Apart from that, with proper documentation, and if the development expanded to the development of item bank, the instrument can be commercialised and sold to other institutions where similar programme to Diploma in English is offered by the other institutions.



1.13 Summary of the Chapter

This chapter has discussed thoroughly about the Diploma in English Entrance Examination reading and grammar paper, why it is important to develop a proper instrument and the aim of this study. It has listed the objectives as well the research questions that this study intends to answer at the end of this analyses. The chapter proceeds with the descriptions of the research background, the theoretical framework of the study and the operational definitions of the study. The next chapter will discuss further on the review of literature as well as the previous researches.

