



REPUBLIK INDONESIA
KEMENTERIAN HUKUM DAN HAK ASASI MANUSIA

SURAT PENCATATAN CIPTAAN

Menteri Hukum dan Hak Asasi Manusia Republik Indonesia, berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta yaitu Undang-Undang tentang perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra (tidak melindungi hak kekayaan intelektual lainnya), dengan ini menerangkan bahwa hal-hal tersebut di bawah ini telah tercatat dalam Daftar Umum Ciptaan:

- I. Nomor dan tanggal permohonan : EC00201701407, 22 Mei 2017
- II. Pencipta
Nama : **Khairiah Syahabuddin, MHSc.ESL., M.TESOL., Ph.D**
Alamat : Jl. Got-Got No. 2, Sektor Utara, Kopelma Darussalam, Banda Aceh, DI ACEH, 23111
Kewarganegaraan : Indonesia
- III. Pemegang Hak Cipta
Nama : **Khairiah Syahabuddin, MHSc.ESL., M.TESOL., Ph.D**
Alamat : Jl. Got-Got No. 2, Sektor Utara, Kopelma Darussalam, Banda Aceh, DI ACEH, 23111
Kewarganegaraan : Indonesia
- IV. Jenis Ciptaan : Karya Tulis
- V. Judul Ciptaan : **Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia**
- VI. Tanggal dan tempat diumumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indonesia : 12 Desember 2013, di Perth
- VII. Jangka waktu perlindungan : Berlaku selama hidup Pencipta dan terus berlangsung selama 70 (tujuh puluh) tahun setelah Pencipta meninggal dunia, terhitung mulai tanggal 1 Januari tahun berikutnya.
- VIII. Nomor pencatatan : 02330

Pencatatan Ciptaan atau produk Hak Terkait dalam Daftar Umum Ciptaan bukan merupakan pengesahan atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang dicatat. Menteri tidak bertanggung jawab atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang terdaftar. (Pasal 72 dan Penjelasan Pasal 72 Undang-undang Nomor 28 Tahun 2014 Tentang Hak Cipta)

a.n. MENTERI HUKUM DAN HAK ASASI MANUSIA
REPUBLIK INDONESIA
DIREKTUR JENDERAL KEKAYAAN INTELEKTUAL
u.b.
DIREKTUR HAK CIPTA DAN DESAIN INDUSTRI

Dr. Dra. Erni Widhyastari, Apt., M.Si.
NIP. 196003181991032001

LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW

HKI

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
Penulis : Khairiah Syahabuddin
Pengusul : Khairiah Syahabuddin
Status Pengusul : Pengusul tunggal
Identitas HKI : a. No. Sertifikat : 02330
b. No. dan Tanggal Permohonan : EC00201701407, 22 Mei 2017
c. Penerbit : Kemenkumham RI
d. Tahun : 2017
e. Link Penerbit : <https://pdki-indonesia.dgip.go.id/detail/EC00201701407?type=copyright&keyword=Achievement%2C+attitude%2C+and+behavior+in+bilingual>

Kategori : Sertifikat HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
Sertifikat HKI, Nomor Pencatatan 02330	10	10
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

Sangat bagus dan sesuai

Banda Aceh, 21 Februari 2022
Reviewer 1,


Prof. Dr. T. Zulfikar, M.Ed.

NIP. 197804302001121002

Guru Besar/Profesor Bidang English Language Research pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

**LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW
HKI**

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
Penulis : Khairiah Syahabuddin
Pengusul : Khairiah Syahabuddin
Status Pengusul : Pengusul tunggal
Identitas HKI : a. Penerbit : Kemenkumham RI
b. Tahun : 2017

Kategori : Konten HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
a. Kelengkapan unsur isi HKI (10%)	1	1
b. Ruang lingkup dan kedalaman pembahasan (30%)	3	3
c. Kecukupan dan kemutakhiran data/informasi dan metodologi (30%)	3	3
d. Kelengkapan unsur dan kualitas penerbit (30%)	3	3
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

a. Kelengkapan unsur isi HKI:

unsur isi lengkap

b. Ruang lingkup dan kedalaman pembahasan:

Ruang lingkup dan pembahasan mendalam

c. Kecukupan dan kemutakhiran data/informasi dan metodologi:

Data dan metodologi mutakhir

d. Kelengkapan unsur dan kualitas penerbit:

Berkualitas

Banda Aceh, 21 Februari 2022
Reviewer 1,


Prof. Dr. T. Zulfikar, M.Ed.

NIP. 197804302001121002

Guru Besar/Profesor Bidang English Language Research pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

LINK DISERTASI DARI EDITH COWAN UNIVERSITY
(ECU) AUSTRALIA:

<https://ro.ecu.edu.au/theses/1083/>

2013

Student English achievement, attitude and behaviour in bilingual and monolingual schools in Aceh, Indonesia

Khairiah Syahabuddin
Edith Cowan University

Recommended Citation

Syahabuddin, K. (2013). *Student English achievement, attitude and behaviour in bilingual and monolingual schools in Aceh, Indonesia*. Retrieved from <https://ro.ecu.edu.au/theses/1083>

This Thesis is posted at Research Online.
<https://ro.ecu.edu.au/theses/1083>

Theses

Theses: Doctorates and Masters

Edith Cowan University

Year 2013

Student English achievement, attitude
and behaviour in bilingual and
monolingual schools in Aceh, Indonesia

Khairiah Syahabuddin
Edith Cowan University, ksyahabu@our.ecu.edu.au

This paper is posted at Research Online.
<http://ro.ecu.edu.au/theses/1083>

Edith Cowan University

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Student English Achievement,
Attitude and Behaviour
in Bilingual and Monolingual Schools
in Aceh, Indonesia

Khairiah Syahabuddin

S. Ag. (Bachelor of Arts in Islamic Studies), State Institute for Islamic Studies
(IAIN Ar-Raniry), Banda Aceh, Indonesia
M.H.Sc. in ESL, International Islamic University, Malaysia
M. TESOL, University of Southern Queensland, Australia

This dissertation is submitted for the degree of
Doctor of Philosophy

School of Education, Faculty of Education and Arts
Edith Cowan University, Perth, Western Australia

December 2013

Copyright Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

- (i) incorporate without acknowledgement material previously submitted for a degree or diploma in any institution of higher education;
- (ii) contain any material previously published or written by another person except where due reference is made in the text; or
- (iii) contain any defamatory material.

I also grant permission for the library of Edith Cowan University to make duplicate copies of my thesis as required.

Signed

Khairiah Syahabuddin

Date: December 11, 2013

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

Motto

من عرف لغة قوم سلم من مكرهم

In Bahasa Indonesia:

*“Barang siapa menguasai bahasa suatu kaum,
maka ia akan selamat dari tipu daya mereka.”*

*(Sebagai syarah [penjelasan singkat] dari Hadits Rasulullah [shallallaahu alaihi
wasallam] tentang kisah Zaid bin Tsabit agar mempelajari bahasa Suryani
[bahasa kaum Yahudi])*

In Acehnese:

*“Baranggasöe carông basa sabôh sabôh kawôm,
maka ureungnyan jeut seulamat dari tipee ngôn muslihat kawôm nyan.”*

*(Sebagöe syarah dari Hadist Rasulullah [shallallaahu alaihi wasallam] teuntang kisah
Zaid bin Tsabit nak geumurunöe bahsa Suryani [bahsa kawôm Yahudi])*

In English:

*“Those who master the language of a people,
will be safe from their conspiracy.”*

*(As a short explanation from Hadits of Rasulullah [peace upon him] on the experience
of Zaid bin Tsabit in order for him to learn the Suryani language [Jewish])*

Abstract

Following the tsunami in 2004, the education system in Banda Aceh, Indonesia, was reconstructed and revitalised, and part of this involved foreign intervention in setting up bilingual schools alongside state-run monolingual schools. The purpose of this study is threefold. The first is to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English essay writing, English reading comprehension, and attitude and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual and monolingual schools). The second is to investigate attitude and behaviour of students with regard to the learning of English as a foreign language, especially regarding student ability in English. The third is to explore students' beliefs and perceptions regarding their experiences of learning English as a foreign language.

A number of linear unidimensional scales were created for each of the three variables using Rasch Measurement with the 2010 RUMM computer program. The construct validity of the three variables was tested by designing the items in ordered patterns of item difficulty which were compared with their Rasch-measured item difficulties, as a Science-like test of the structure of the variables. An experimental research design (pretest/posttest, control/experimental group) was used with Rasch-created linear measures of three variables: (1) a researcher-designed English Essay Test; (2) a researcher-designed Reading Comprehension Test; and (3) a researcher-designed Attitude/Behaviour Test about Learning English. Seven hundred and eighty male and female first-year middle school students (aged 12-13 years old), consisting of 394 students from bilingual schools and 386 students from monolingual schools, selected from a number of schools with bilingual programs and monolingual programs, were the respondents for this study. After two months of lessons, the two groups were compared on each of the three measures using ANCOVA and ANOVA. Students' written comments were collected in regards to their experiences of learning English as a foreign language.

The findings showed that bilingual students outperformed monolingual students in tests of English Reading Comprehension, English Writing and Attitude/Behaviour for both pretests and posttests. Female students achieved better results than male students in English Reading Comprehension, English Writing, and Attitude/Behaviour tests, both for pretests and posttest.

Research Papers Based on this Dissertation

Syhabuddin, K. & Waugh, R. F. (2012a). Differential item functioning of English reading comprehension in a second language context when fit to the Rasch model is not ideal [Peer-reviewed]. Paper presented at the Fifth International Conference on Probabilistic Models for Measurement, held at the University of Western Australia, Perth, 23-25 January 2012.

Syhabuddin, K. & Waugh, R. F. (2012b). Students' ability in English reading comprehension by gender and bilingual/monolingual learning: An Acehese case (Banda Aceh). Paper presented at the Western Australia Institute for Education Research (WAIER), held at Notre Dame University, Fremantle, 11 August 2012.

Syhabuddin, K. & Waugh, R. F. (2012c). Students' Behaviour towards learning English as a foreign language by gender and bilingual/monolingual learning: An Acehese case (Banda Aceh). Paper presented at the Applied Linguistics Association of Australia (ALAA) National Conference, held at Curtin University, Perth, 12-14 November 2012.

<http://humanities.curtin.edu.au/schools/EDU/education/alaa-proceedings.cfm>

Acknowledgement



In the Name of Allāh, the Most Gracious, the Most Merciful

Three years ago, at the beginning of 2010, when starting this research journey, I was picturing myself as a person trying to reach the peak of a mountain with the help of a thin rope. Not easy. Being a full-time student while raising and nurturing two little children, aged 8 and 2.5 then, is not easy. The completion of this dissertation, therefore, is possible because of the abundant help and support I received. I am so thankful.

The most and foremost thanks, *Alhamdulillah* *‘Ālamēn*, goes to my God, Allah *Subhānahuwa Ta’ālā*, the Lord of the Universe. It is because of His Love I managed to finalise this dissertation. Grand blessings and peace are upon Prophet Muhammad (peace upon him) who has brought mankind from the darkness to the lightness of religion and knowledge. I am one of many who is given an opportunity to enrich my religion and knowledge.

My greatest appreciation goes to Professor Russell Waugh, my principal supervisor, who accepted me as one of his postgraduate students. It is he who introduced me to Rasch Measurement, a powerful research breakthrough, which I did not know before. He taught and nurtured me from the very beginning until this dissertation was completed. That involves thinking, reading, discussions, writing drafts, revising and re-revising. He makes sure I am ‘on track’, with, “*Khairiah, when are you going to give a new chapter to Dr Russell?*” or “*Well, Khairiah, is this the week you have a new chapter ready for me?*” When drafts are ready for him to read, it is only one or two days waiting, for the drafts return quickly to me. A lightning supervisor he is. I am deeply grateful. My appreciation also goes to my co-supervisor, Dr Anne Thwaite, who helped me with interesting discussions and thorough corrections.

My great thanks are to ECU, its School of Education, Graduate Research Support, and SOAR. My sincere thanks to the ECU Research Writing Consultant, Dr Jo McFarlane, and to my ECU roommates at Mount Lawley Postgraduate Room 16.238: Sulfasyah Syaharuddin (Sulfa), Omolola Aina Ladele (Lola), and Steven Chia (Steve) who share tips and tricks regarding writing up a dissertation.

I am also blessed by having friends from Indonesia and my hometown, Aceh, who helped cheer me up through BBQ gatherings, park picnics, Moslem celebrations, chit-chats and others who made me feel that I was not alone. They were beside me to offer help.

I also thank my home university, Ar-Raniry State Islamic University, Banda Aceh who released me on my study leave to pursue this PhD. Also my gratitude goes to 18 State Middle Schools in Banda Aceh who participated in this study: their headmasters, English teachers, and students.

My deepest gratitude goes to the Australia Endeavour Scholarship which supported my study for four years.

My warmest thanks go to my family in Banda Aceh, the family of *Bapak* Syahabuddin Mahyiddin: My mother *Mamak* Cut Nurhayati, H. Ahmad; my older brothers: *Bang* Rusydi Syahabuddin, *Bang* Mursyid Syahabuddin, *Bang* Armia Syahabuddin, and my only and younger sister, *Adek* Fadhilah Syahabuddin. Thanks also are to my sisters and brother-in-law, and cousins. My sincere thanks for them all who have always prayed and supported me with anything that one can imagine.

The last and the highest thanks goes to my family: My husband, *Kakanda* Sa'i Baharuddin; my daughter, Farhanah Arifah Sa'i; and my son, Tijan al-Darary Sa'i, who allowed me to take this study in the first place and to experience life in Perth; who were with me in my highs and lows; and who agreed to share some of their weekends, public holidays and school holidays with my work on this dissertation.

I am deeply thankful.

May we all be blessed.

Dedication

This dissertation is dedicated to my sunshine

daughter, **Farhanah Arifah Sa'i**,
and son, **Tijan al-Darary Sa'i**,
who cherish their life by being bilinguals

And to my husband, **Kakanda Sa'i Baharuddin**,
who is starting to be a bilingual

Also, to my late father, **Bapak Syahabuddin Mahyiddin**,
and my mother, **Mamak Cut Nurhayati H. Ahmad**,
who encouraged me to be a bilingual

Table of Contents

Front Page	i
Copyright Declaration.....	iii
Use of Thesis.....	v
Motto.....	vii
Abstract	ix
Research Papers Based on this Dissertation	xi
Acknowledgement	xiii
Dedication	xv
Table of Contents	xvi
List of Tables	xxvi
List of Figures	xxx
List of Appendices	xxxiv
CHAPTER ONE	1
INTRODUCTION.....	1
Background	1
Aceh Province	1
Tsunami Aftermath	2
Education in Aceh Province.....	4
Bilingual and Monolingual Education	10
Rationale	12
Statement of Purpose	13
Research Questions	13

Significance.....	14
Limitations	15
Definition of Terms.....	17
Structure of the Thesis	23
CHAPTER TWO	25
BILINGUALISM and BILINGUAL EDUCATION	25
The Nature of Bilingualism.....	25
What does it mean to be a bilingual?.....	25
What are the Benefits of Being a Bilingual?	31
Is there any Disadvantage to being Bilingual?	34
The Nature of Bilingual Education	36
What is Bilingual Education?	36
What are the Existing Bilingual Education Models and Practices?	37
Brief History of Bilingual Education Worldwide.....	45
The Nature of Bilingual Education in Indonesia.....	50
What is the Position of Bahasa Indonesia?.....	50
Why is English in the Indonesian Curriculum?.....	50
Why Do They Have Bilingual Education in Indonesia?	51
Bilingual Education Program Ban in Indonesia (2013).....	52
What is the future of Bilingual Education in Indonesia?.....	53
A Review of Bilingualism and English as a Second or Foreign Language	53
Review of Second Language Acquisition.....	54
Findings on Experimental and Control Studies for Reading and Writing, and Attitude and Behaviour	58
Findings Using Rasch Measures and Control and Experimental Groups.....	63

Gaps in Indonesian Research: Bilingualism <i>versus</i> Monolingualism.....	64
CHAPTER THREE	67
CONCEPTUAL FRAMEWORK FOR BILINGUAL/MONOLINGUAL TEACHING, MEASUREMENT, AND THE VARIABLES	67
Conceptual Framework for Bilingual and Monolingual Teaching in Banda Aceh	67
Measurement	71
True Score Theory Measurement	71
Rasch Measurement.....	72
Variables	78
Measuring English Reading Comprehension	78
Measuring English Writing Achievement	80
Questionnaire on Attitude and Behaviour with regard to Learning English	83
CHAPTER FOUR	85
RESEARCH DESIGN	85
Strategy and Design	85
Mixed-Method Design	86
Samples	89
Instruments for English Tests	92
Questionnaire on Student Attitude/Behaviour	93
Instruments for students' written comments	94
Administration of the Instruments	94
Procedure	94
Pilot Studies.....	95
Ethical Considerations and Study Approval	100

Ethics Approvals.....	100
Study Approval.....	102
Control of Extraneous Variables in the Quasi-Experiments	102
Test Data Collection, and Data Entry	104
Test Data Collection	104
Data Entry.....	105
Data Analysis	105
Analysis of Reading Comprehension and Essay Writing Test data	105
Analysis of Attitude and Behaviour Questionnaire Data	106
Analysis of Students' Written Comments	107
CHAPTER FIVE.....	109
DATA ANALYSIS (PART 1).....	109
RASCH MEASUREMENT OF READING COMPREHENSION.....	109
Output Analysis.....	110
Standardised Fit Residuals.....	110
Principal component analysis of residuals.....	111
Item-trait interaction (dimensionality).....	111
Person Separation Index	112
Individual Item Fit	113
Threshold Values	114
Scoring Category Curve	114
Item Characteristic Curves	116
Person-Item Threshold Distribution (Targeting).....	118
Differential Item Functioning (DIF) by Gender	119
Differential Item Functioning (DIF) by Type of Language Instruction	121

Targeting by Gender and Type of Language Instruction	123
Scale of Item Difficulties.....	124
Discussion and Summary.....	125
CHAPTER SIX	129
DATA ANALYSIS (PART 2).....	129
EXPERIMENTAL RESULTS FOR READING COMPREHENSION	129
Results on Mixed Design ANOVA (Split-Plot Design ANOVA).....	129
Results on Two-Way ANCOVA.....	133
Summary	135
CHAPTER SEVEN.....	139
DATA ANALYSIS (PART 3).....	139
RASCH MEASUREMENT OF ENGLISH WRITING	139
Initial Analysis (Posttest).....	139
Output from the Final Rasch Analysis (Postest).....	144
Standardised Fit Residuals.....	144
Dimensionality and Item-Trait Interaction	145
Reliability	146
Item Statistics	146
Scoring Category Curve	147
Item Characteristic Curves	148
Person-Item Threshold Distribution (Targeting).....	150
DIF by Gender	151
DIF by Type of Language Instruction	154
Targeting by Gender and Type of Language Instruction	154

Scale of Item Difficulties.....	156
Pretest.....	156
Discussion and Summary.....	157
CHAPTER EIGHT	159
DATA ANALYSIS (PART 4).....	159
EXPERIMENTAL RESULTS FOR ENGLISH WRITING	159
Results for Mixed Design ANOVA (Split-Plot Design ANOVA)	160
Results for Two-Way ANCOVA.....	163
Summary	165
CHAPTER NINE	169
DATA ANALYSIS (PART 5).....	169
RASCH MEASUREMENT OF ATTITUDE/BEHAVIOUR	169
Initial Analysis	169
Output from Final Analysis.....	170
Standardised Fit Residuals.....	170
Dimensionality and Item-Trait Interaction	170
Person Separation Index	171
Individual Item Fit	171
Threshold Values	172
Scoring Category Curve	173
Item Characteristic Curves	174
Person-Item Threshold Distribution (Targeting).....	175
DIF by Gender	177
DIF by Type of Language Instruction	179

Targeting by Gender and Type of Language Instruction	181
Scale of Item Difficulties.....	183
Discussion and Summary.....	183
CHAPTER TEN.....	185
DATA ANALYSIS (PART 6).....	185
EXPERIMENTAL RESULTS FOR BEHAVIOUR.....	185
Results for Mixed Design ANOVA (Split-Plot Design ANOVA)	185
Results for Two-Way ANCOVA.....	189
Summary	191
CHAPTER ELEVEN.....	193
DATA ANALYSIS (PART 7).....	193
STUDENTS' WRITTEN COMMENTS.....	193
Theme 1: Tasks on Learning English as a foreign language	194
Tasks for Listening	194
Tasks for Speaking	195
Tasks for Reading.....	195
Tasks for Writing.....	195
Theme 2: Student-Student Relationships.....	196
Theme 3: Student-Teacher Relationships	197
Like the way English is taught	197
Dislike the way English is taught	198
Theme 4: Personal Views on the Benefits of Learning English	198
Theme 5: Common Views on the Benefits of Learning English	201
Theme 6: Confidence and Achievement in Learning English	203

Theme 7: Learning English Through the Media	205
Theme 8: Family Support in Learning English.....	206
Theme 9: Obstacles in Learning English	208
Theme 10: Other Views	210
Summary	211
CHAPTER TWELVE.....	213
DISCUSSION OF RESULTS AND IMPLICATIONS.....	213
Summary of the Present Study	214
Answering the Research Questions.....	215
Summary of the Findings	220
A. English Reading Comprehension	220
B. English Writing	221
C. Behaviour with regard to Learning English as a Foreign Language.....	221
Discussion about Bilingual Education	222
Implications for Policy and Practice	226
Implications for Policy: Indonesian Ministry of Education/Policy Makers	227
Implications for Practice.....	227
Implication for Classroom Practices.....	230
Implications for Parents	231
Implications for Further Research	231
Limitations of the Study	233
Managing Time and Extraneous issues	233
Managing Teacher Perceptions (through Interviews).....	235
Managing Classroom Observations.....	235
Measures were not Ideal and Need Improvement	235

REFERENCES	237
APPENDICES	251
Appendix A-English Reading Comprehension.....	251
Appendix B-English Reading Comprehension’s Ordered Scoring Scheme.....	255
Appendix C-English Writing Test	256
Appendix D-English Writing Scoring Rubric.....	258
Appendix E-Attitude/Behaviour Questionnaire.....	260
Appendix F-Original English Reading Comprehension Text.....	262
Appendix G-Original Attitude/Behaviour Test.....	264
Appendix H-Request Letter to Head of Provincial Education of Aceh.....	265
Appendix I-Consent Letter to Headmaster	266
Appendix J-Consent Letter to English Teacher	267
Appendix K-Consent Letter to Student (English Reading Comprehension)	268
Appendix L-Consent Letter to Student (English Writing).....	269
Appendix M-Consent Letter to Student (Attitude/Behaviour Questionnaire)	270
Appendix N-Student’s Sample Answer of Reading Comprehension	271
Appendix O-Student’s Sample Answer English Writing (Bilingual Student).....	272
Appendix P-Student’s Sample Answer English Writing (Monolingual Student).....	274
Appendix Q-Student’s Sample Answer Attitude/Behaviour Questionnaire (Bilingual Student).....	276
Appendix R-Student’s Sample Answer Attitude/Behaviour Questionnaire (Monolingual Student).....	277
Appendix S-Complete Occurrence of Student Written Comment.....	278

List of Tables

Table 1.1 <i>Formal Education of Indonesia</i>	6
Table 2.1 <i>Typology of Bilingualism Classical</i>	29
Table 2.2 <i>Type of Bilingual Education</i>	44
Table 2.3 <i>Brief History of Bilingual Education in the USA</i>	46
Table 2.4 <i>Difference Among the Four Perspectives</i>	54
Table 3.1 <i>Item Difficulty Scoring Scheme for Reading Comprehension Test</i>	79
Table 3.2 <i>English Writing Scoring Rubric</i>	82
Table 3.3 <i>Attitude and Behaviour Questionnaire</i>	84
Table 4.1 <i>Student Numbers by Control/Experimental Groups and by Gender in Control/Experimental Groups</i>	91
Table 5.1 <i>Overall Fit Statistics Reading Comprehension Measure (N=780, I=12)</i>	111
Table 5.2 <i>Item-Trait Interaction for Reading Comprehension Scale</i>	112
Table 5.3 <i>Locations, Standard Errors, Residuals and Chi-Squares for Reading Comprehension Items</i>	113
Table 5.4 <i>Item Thresholds-uncentralised (Item=12, Number=780) for Reading Comprehension Measure</i>	114

Table 5.5	
<i>Order of Difficulty of Items on the Linear Scale</i>	124
Table 6.1	
<i>Pretest and Posttest Scores for the Bilinguals and Monolinguals</i>	132
Table 6.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	132
Table 6.3	
<i>Interaction Effect and Main Effect of Reading Comprehension between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	134
Table 7.1	
<i>Writing Marking Example for topic 1 ‘My Idul Fitri Holiday’</i>	142
Table 7.2	
<i>Initial Marking of English Writing Essay as Items (Scored in terms of 18 Items for Compatibility with RUMM2030)</i>	143
Table 7.3	
<i>Final Marking of English Writing Essay as Items (Scored in terms of 12 Items for Compatibility with RUMM2030)</i>	144
Table 7.4	
<i>Overall Fit Statistics for the Writing Test (N=780, I=12)</i>	145
Table 7.5	
<i>Item-Trait Interaction for the Writing Scale</i>	146
Table 7.6	
<i>Locations, Standard Errors, Residuals and Chi-Squares for Writing Items</i>	171
Table 7.7	
<i>Order of Difficulty of Items on the Linear Scale</i>	156
Table 8.1	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	161
Table 8.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	162
Table 8.3	
<i>Interaction Effect and Main Effect of Writing between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	163

Table 9.1	
<i>Overall Fit Statistics for the Behaviour Measure (N=779, I=11)</i>	170
Table 9.2	
<i>Item-Trait Interaction for Behaviour Scale</i>	171
Table 9.3	
<i>Locations, Standard Errors, Residuals and Chi-Squares for Behaviour Items</i>	172
Table 9.4	
<i>Item Thresholds Uncentralised (Item=11, Number=779) for Behaviour Measure</i>	173
Table 9.5	
<i>Order of Difficulty of Items on the Linear Scale</i>	183
Table 10.1	
<i>Pretest and Posttest Scores for the Bilinguals and Monolinguals</i>	187
Table 10.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	188
Table 10.3	
<i>Interaction Effect and Main Effect of Behaviour between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	189
Table 11.1	
<i>Students' Personal Views on the Benefits of Learning English as a Foreign Language</i>	199

List of Figures

<i>Figure 1.1</i>	The Location of Aceh Province within Indonesia	1
<i>Figure 1.2</i>	A Map of Aceh Province	2
<i>Figure 1.3</i>	An Ogive Curve	22
<i>Figure 2.1</i>	A Mythical View of the 'Ideal' or 'Full' bilingual	26
<i>Figure 2.2</i>	More realistic Conceptions of Bilingualism	26
<i>Figure 4.1</i>	Method Design Chart	88
<i>Figure 4.2</i>	Original Reading Comprehension Test Front Page	97
<i>Figure 4.3</i>	Revised Reading Comprehension Test Front Page	98
<i>Figure 5.1</i>	Item Category Curve for Item 1	115
<i>Figure 5.2</i>	Item Category Curve for Item 6	116
<i>Figure 5.3</i>	Item Characteristic Curve for Item 1	117
<i>Figure 5.4</i>	Item Characteristic Curve for Item 6	117
<i>Figure 5.5</i>	Person-Item Threshold Distribution for Reading Comprehension	118
<i>Figure 5.6</i>	Students' Reading Comprehension and Item Difficulties Calibrated on the Same Scale	119
<i>Figure 5.7</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 10	120
<i>Figure 5.8</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 12	120
<i>Figure 5.9</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 1	121
<i>Figure 5.10</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 2	122
<i>Figure 5.11</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 8	122
<i>Figure 5.12</i>	Targeting of Reading Comprehension by Gender	123
<i>Figure 5.13</i>	Targeting of Reading Comprehension by Type of Language Instruction	124
<i>Figure 6.1</i>	Graph of Reading Comprehension by Group	131

<i>Figure 6.2</i>	Graph of Reading Comprehension by Gender	132
<i>Figure 6.3</i>	Graph of Reading Comprehension by Group (Bilingual/Monolingual) with Pretest as Covariate	134
<i>Figure 6.4</i>	Graph of Reading Comprehension by Gender with Pretest as Covariate	135
<i>Figure 7.1</i>	Curve for Item Category for Item 11	148
<i>Figure 7.2</i>	Item Characteristic Curve for Item 3	149
<i>Figure 7.3</i>	Item Characteristic Curve for Item 9 (Not Ideal but Satisfactory)	149
<i>Figure 7.4</i>	Targeting Graph: Student Measures in English Writing (upper-side) and Item Difficulties (lower-side) Calibrated on the Same Scale	150
<i>Figure 7.5</i>	Writing Map	151
<i>Figure 7.6</i>	Item Characteristic Curve by Gender for Writing Item 12	152
<i>Figure 7.7</i>	Item Characteristic Curve by Gender for Writing Item 3	152
<i>Figure 7.8</i>	Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 12	153
<i>Figure 7.9</i>	item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 3	154
<i>Figure 7.10</i>	Targeting of Writing by Gender	155
<i>Figure 7.11</i>	Targeting of Writing by Type of Language Instruction	155
<i>Figure 8.1</i>	Graph of Writing Test by Group	161
<i>Figure 8.2</i>	Graph of Writing Test by Gender	162
<i>Figure 8.3</i>	Graph of Writing Test by Group (Bilingual/monolingual) with Pretest as Covariate	164
<i>Figure 8.4</i>	Graph of Writing Test by Gender with Pretest as Covariate	165
<i>Figure 9.1</i>	Item Category Curve for Item 2	174
<i>Figure 9.2</i>	Item Characteristic Curve for Item 2	175
<i>Figure 9.3</i>	Person-Item Threshold Distribution for Behaviour	176
<i>Figure 9.4</i>	Behaviour Map	177
<i>Figure 9.5</i>	Item Characteristic Curves by Gender for Behaviour Item 26	178
<i>Figure 9.6</i>	Item Characteristic Curves by Gender for Behaviour Item 12	178
<i>Figure 9.7</i>	Item Characteristic Curves by Gender for Behaviour Item 38	179

<i>Figure 9.8</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 7.....	180
<i>Figure 9.9</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 2	180
<i>Figure 9.10</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 38	181
<i>Figure 9.11</i>	Targeting of Behaviour by Gender	182
<i>Figure 9.12</i>	Targeting of Behaviour by Type of Language Instruction	182
<i>Figure 10.1</i>	Graph of Behaviour by Group	187
<i>Figure 10.2</i>	Graph of Behaviour by Gender	188
<i>Figure 10.3</i>	Graph of Behaviour by Group (Bilingual/monolingual) with Pretest as Covariate	190
<i>Figure 10.4</i>	Graph of Behaviour by Gender with Pretest as Covariate	191
<i>Figure 11.1</i>	Students' Common Views on the Benefits of Learning English	201
<i>Figure 11.2</i>	Students' Confidence and Achievement.....	204
<i>Figure 11.3</i>	Learning English through the Media.....	205
<i>Figure 11.4</i>	Family Support in Learning English	207
<i>Figure 11.5</i>	Obstacles in Learning English	209
<i>Figure 11.6</i>	Other Views.....	210
<i>Figure 12.1</i>	Bilingualism as a Life Time Skill	226

List of Appendices

Appendix A	English Reading Comprehension Test	251
Appendix B	English Reading Comprehension's Ordered Scoring Scheme	255
Appendix C	English Writing Test	256
Appendix D	English Writing Scoring Rubric	258
Appendix E	Attitude/Behaviour Questionnaire	260
Appendix F	Original English Reading Comprehension Text	262
Appendix G	Original Attitude/Behaviour Test	264
Appendix H	Request Letter to Head of Provincial Education of Aceh	265
Appendix I	Consent Letter to Headmaster	266
Appendix J	Consent Letter to English Teacher	267
Appendix K	Consent Letter to Student (English Reading Comprehension)	268
Appendix L	Consent Letter to Student (English Writing)	269
Appendix M	Consent Letter to Student (Attitude/Behaviour Questionnaire)	270
Appendix N	Student's Sample Answer for Reading Comprehension	271
Appendix O	Student's Sample Answer for English Writing (Bilingual Student)	272
Appendix P	Student's Sample Answer for English Writing (Monolingual Student)	274
Appendix Q	Student's Sample Answer for Attitude/Behaviour Questionnaire (Bilingual Student)	276
Appendix R	Student's Sample Answer for Attitude/Behaviour Questionnaire (Monolingual Student)	277
Appendix S	Complete Students' Written Comments.....	278

CHAPTER ONE

INTRODUCTION

Background

This chapter introduces the reader to the Indonesian education system and to bilingual and monolingual education in Aceh province, Indonesia, and to the rationale and background behind the study, and thus to its aims.

Aceh Province

Aceh (pronounced ä'chā), is located on the northern tip of Sumatra island, bordered on the north by the Malacca Strait and on the south by the Indian Ocean. Its population reached 4,144,500 in 2002, with an area of 57,365.57 square kilometers (Kadin Aceh, nd). The provincial capital is Banda Aceh, which had a population before the tsunami of 264,618 and after the tsunami 203,553 (Nurdin, 2006). The first part of its name comes from the Persian *bandar* and means "port" or "haven" (Washington, 2010). Banda Aceh is known as 'Mecca's Verandah' (or 'doorway to Mecca'), because historically it has been a stopping place for Muslim pilgrims journeying by ship from the east to Mecca and as Islam entered the Indonesian Archipelago through Aceh. Figure 1.1 shows the location of Aceh Province within Indonesia, and Figure 1.2 is a map of Aceh Province.



Figure 1.1 *The Location of Aceh Province within Indonesia*



Figure 1.2 *The Map of Aceh Province*

Having been previously given five different names: *Aceh Lhee Sagoe*, *Seuramoe Makkah*, *Nangroe Aceh Darussalam*, *Banda(r) Aceh* and *Kutaraja*, Aceh province is rich in resources, mainly oil and natural gas. The most important agricultural products are rice, coconuts, rubber, tea and coffee. Due to its location, Aceh has been one of the first points of contact for foreigners in Sumatra. Indian, European and Arabic visitors have all had great influence through the spreading of religion and culture. Over different periods of time, its location by the Malacca strait has also led to intensive trade and high economic growth. Buddhism came to Aceh with Indian traders in the 7th century, while Islam was introduced by Arabic and Indian Muslims in the 13th century. A number of Muslim kingdoms and sultanates were subsequently established in the region. Aceh (formerly called Achin) reached the peak of its power in the early 17th century. The Dutch gained control of the coast in 1873 and engaged in a partly successful effort to subdue the interior until 1910 (Answer.Com, 2010). The Acehnese, like most Indonesians, are Muslim, but are generally more conservative. Aceh is today considered as the province in Indonesia where Islam has the strongest position.

Tsunami Aftermath

Until December 26, 2004, Banda Aceh was not frequently the subject of international discussion. On that day, a giant earthquake of 9.3 on the Richter scale occurred off the coast of Aceh province. The earthquake started a huge tsunami, sending massive tidal waves towards the coast, which destroyed a large number of coastal areas

in Aceh, Indonesia and numerous islands in the province of North Sumatra. The tsunami buried parts of the cities of Meulaboh and Banda Aceh (Community webs, 2010). It was the worst hit area out of all the locations hit: 220,000 people were estimated to have died in January 2005; and by 2009, 225,000 were estimated to have died (Education, 2010). According to the country's National Disaster Relief Coordination Agency, another 37,063 are still missing. In addition, the UN estimates that 655,000 people were homeless and sheltering in scattered refugee camps across the province (BBC, 2005). The highest tolls were among the women and children who were in the low lying coastal areas while their husbands were at sea fishing. Over 800 kilometres of coast were severely affected, often up to five kilometres inland. At least 654 villages were damaged or destroyed, more than 500,000 people lost their homes, and more than 150,000 children were left without schools (Education, 2010).

The Indonesian government, supported by massive financing provided by national and international sources, responded quickly, providing food, personnel and aid. In the first days, weeks and months following the disaster they were focused on providing emergency relief necessities such as food, clean water and shelter, and the resources needed to treat victims, stop the spread of disease, clear debris from roads and restore essential services. The swift and well-coordinated emergency effort managed to prevent a major outbreak of disease and hunger (ADB, 2009; Education, 2010).

Reconstruction and rehabilitation in Aceh were based on emergency relief demands and implemented with money, goods and services provided by the Indonesian government, the United Nations and extensive international support. This has resulted in major results and improvements. Total commitment, as of 31 December 2008, was \$6.8 billion and this has exceeded the minimum estimated to re-build to pre-tsunami levels (estimated at \$6.2 billion) (ADB, 2009). As of December 2008, a \$2.4 billion fund was used for housing and transport, and \$1.7 billion was used for health, education, and community-based development activities. Regarding the education sector, 395 schools covering all 23 districts in Aceh have received support for teaching and learning materials, books, computers, and repairs for libraries (ADB, 2009).

Education in Aceh Province

In reconstruction and rehabilitation in the educational sector, the Acehnese have decided to proceed with an Indonesian policy. As part of Indonesia, the Aceh government is entitled to run a national policy. The difference between the education systems in Aceh province from that of other provinces is due to the education privilege given to Aceh, in that Islamic wisdoms are added as local content for primary and secondary education.

The National Education System of Indonesia is generally aimed at elevating the intellectual life of the nation and developing the Indonesian people fully, that is, as people who are devoted to God, have good knowledge and high levels of skill, are in good physical and spiritual health, are independent and fair, and feel responsible for their countrymen and nation. The education system also strives to create patriotic spirit, strengthen love for the fatherland, enhance social solidarity and awareness of the nation's history regarding its heroes, and create a forward-looking attitude. The learning and teaching climate has to generate self-confidence and a learning culture among all layers of society that induces an attitude and behaviour of creativity, innovative thinking, and orientation toward the future. The education system has its roots in the Indonesian culture, based on *Pancasila* (the five official philosophical foundations of Indonesia) and the 1945 Constitution. Law No. 2/1989 concerns the National Education System, the system aims at generating abilities and increasing the standard of living and dignity of the Indonesian people in order to achieve the national development objectives (Menanet, n.d).

The entire national curriculum of Indonesia is designed around two basic parts, the Pancasila and the 1945 Constitution. Based on the Indonesian constitution, the goal of education is to prepare its citizens (and students) to develop creatively and emotionally and acquire the skills, knowledge, attitudes and spiritual values necessary for responsible, active, autonomous, productive, and democratic citizenship (Syahdan, 2009). The constitution also notes that education in Indonesia is divided into two major parts, formal and non-formal. A formal education is standard education obtained through an accredited source which is divided into three levels, primary, secondary and tertiary education (see Table 1.1). A non-formal education is usually run at pre-school

age and primary level, such as TPA abbreviation for Taman Pendidikan Al-Quran (Centre for Al-Quran Learning) which is usually held in mosques, Sunday schools which are usually held by churches, and music tuition, English tuition, archery classes, and so on that are held by various providers.

Primary education consists of early childhood and elementary school. From birth until the age of three, Indonesian children do not generally have access to formal education. From the age of three to four or five, they attend kindergarten (*Taman Kanak-kanak*). This education is not compulsory for Indonesian citizens, as the aim of kindergarten is to prepare them for primary school. The majority of kindergartens are private schools and 99.35% of the kindergartens (i.e. 49,000) in Indonesia are privately operated. The kindergarten years are usually divided into "Class A" and "Class B", students spending a year in each class. Children aged 6–12 attend *Sekolah Dasar* (SD) (literally Elementary School). This level of education is compulsory for all Indonesian citizens, based on the national constitution. In contrast to the majority of privately run kindergartens, most elementary schools are government-operated public schools, accounting for 93% of all elementary schools in Indonesia. Similar to education systems in the U.S.A and Australia, students must study for six years to complete this level. Some schools offer an accelerated learning program, where students who perform well can finish elementary school in five years (Exchange., n.d)

Secondary Education in Aceh consists of middle school and high school. Middle School, generally known by the abbreviation "SMP" (*Sekolah Menengah Pertama*) is part of primary education in Indonesia. Alternatively students can enroll in *Madrasah Tsanawiyah*, abbreviated to "MTs". SMP and MTs have a slightly different curriculum; with MTs has additional subjects on Islamic studies. After graduating from elementary school, students attend Middle School for three years from the age of 12-14. After three years of schooling and graduation, students may move onto High School, or cease formal education. There are around 22,000 schools in Indonesia, with a balanced ownership between the public and private sectors. In Indonesia, high school is generally known by the abbreviation 'SMA' (*Sekolah Menengah Atas*) and 'SMU' (*Sekolah Menengah Umum*)/'SMK' (*Sekolah Menengah Kejuruan*). SMA/SMU differs from SMK in its curriculum studies and student destinations. The students at SMU are prepared to advance to tertiary education while students of SMK are prepared to be

ready to join the workforce after finishing their school without going to university or college. Alternatively students can enroll in *Madrasah Aliyah*, abbreviated to “MA”. SMA/SMU and MA have a slightly different curriculum; with MA has additional subjects on Islamic studies. Based on the national constitution, Indonesian citizens do not have to attend high school, as the citizens are only required by law to attend school for nine years of education. This is reflected in the relatively low number of high schools in Indonesia, just slightly below 9,000 schools.

Tertiary education occurs after graduation from high school or college. Students may attend a university or higher education institution. Higher education institutions are categorized into two types, public and private, both of which are supervised by the Department of National Education. There are three types of higher education institution: Universities, Institutes and Academies or colleges (Mohandas, n.d).

Table 1.1
Formal Education of Indonesia

LEVEL OF EDUCATION	Year	Age
	Kindergarten	
	Playgroup	4
	Group A	5
	Group B	6
	Primary School (<i>Sekolah Dasar</i> or SD, <i>Madrasah Ibtidaiyah</i> or MI)	
PRIMARY EDUCATION	Year 1	7
	Year 2	8
	Year 3	9
	Year 4	10
	Year 5	11
	Year 6	12
	Middle School (<i>Sekolah Menengah Pertama</i> or SMP, <i>Madrasah Tsanawiyah</i> or MTs)	
SECONDARY EDUCATION	Year 7	13
	Year 8	14
	Year 9	15
	High School / Vocational School (<i>Sekolah Menengah Atas</i> or SMA and <i>Madrasah Aliyah</i> or MA / <i>Sekolah Kejuruan</i> or MK)	

SECONDARY EDUCATION	Year 10	16
	Year 11	17
	Year 12	18
Academy/Institute/Polytechnic/College/ University		
TERTIARY EDUCATION	Bachelor	Various ages (approx. 4 years)
	Master	Various ages (approx. 2 years)
	Doctorate	Various ages (approx. 4 years)

Indonesia has a national education system which controls the education system across the whole archipelago, including Aceh. This system was regulated under Law No.2/1989, and then changed to Law No. 20/2003 (Wijaya, 2008). When the education system was under Law No. 2/1989, the curriculum implemented was Curriculum 1994, which focus was content oriented. The grading system focused more on the cognitive side of student learning. The learners had to study in class for 40 hours every week. The syllabi used were under centralised control and all schools had to use the same syllabi. The focus of English language teaching in Curriculum 1994 was grammar and reading and the teachers were the source of information (Elvyanti, n.d).

After being used for ten years, curriculum 1994 was changed into Curriculum 2004, which was called the competence-based curriculum (*Kurikulum Berbasis Kompetensi* or KBK) (Elvyanti, n.d). Curriculum 2004 was under Law No. 20/2003. In Curriculum 2004, the focuses were product and competence. The concepts of learning were learning to know, learning to do, learning to live together and learning to be. The grading system combined the cognitive side, the affective side and the psychomotor side all together. The learning time was reduced from 40 hours to 32 hours every week, although the number of subjects taught was the same as Curriculum 1994. In the 2004 curriculum, the government gave freedom to teachers to create the syllabi which could fulfill the needs of the local area and the learners (Elvyanti, n.d), including those students in Aceh province.

However, Curriculum 2004 is now not used any longer. It has been changed into Curriculum 2006, but is still the curriculum under Law No. 20/2003, and is a further development of Curriculum 2004. In the implementation of this 2006 curriculum, schools are given the authority to design the syllabus, the learning hours, the academic

calendar and the standard competence of graduates (Elvyanti, n.d). Moreover, there are more parties – societies, parents and school committees – involved in the syllabus design.

Curriculum 2006 has been implemented by primary and secondary levels of education. The procedures of curriculum implementation are regulated under Law No. 20/2003. Meanwhile, the tertiary level of education is to implement Curriculum 2004. The curriculum implementation at the tertiary level is regulated under Law No. 20/2003 and the Decree of Minister of National Education No. 232/U/2000 (Nasional, 2001). This curriculum implementation means there are curriculum differences in the primary, secondary and tertiary levels of education which affect the teaching of English and the English learning process.

Just this year (2013) the new Curriculum 2013 has been implemented. The new feature of the curriculum is the banning of bilingual programs. This means reducing English teaching learning lessons.

In Aceh Province, as well as in other part of Indonesia, English has been taught as a foreign language since 1950 (Yuwono, 2005) and English has been a focus in the curriculum as a compulsory subject from the first year of Middle School (students 12/13 years old) up to the final year of High School (students 17/18 years old), over six years of secondary schools. However, some elementary schools (similar to Primary Schools in the Australian context, with students aged from 6-12), have put English into their school local content. Generally, English is not compulsory in elementary schools.

English is taught as a separate subject and it does not become the medium of instruction at any particular point in schooling. The majority of teachers are Acehnese-Indonesians with diplomas in teaching English as a foreign or second language, and the language is not taught by native speakers or expatriates. English is generally the same way in all schools regardless of the impact of the new curriculum, which now gives the right for schools to design their own content so that it best suits their school's needs.

The idea of bilingualism is not new to the majority of Indonesians. Most of the people in Indonesia are bilingual at an early age. They speak a local native language with their family whereas the official language is Bahasa Indonesia (also known as

Indonesian). As a language of instruction at school, Bahasa Indonesia is often the only language spoken. However, in remote areas throughout Indonesia, teachers need to alternate their teaching with the local native languages. The local native language is not dying out, even though it serves only as a spoken language.

Formal bilingual education is still seen as a ‘fantasy’ and a ‘luxury’ for the majority of Indonesian parents. Those parents who have good incomes mostly enrol their child or children into bilingual schools, which are usually international schools or private schools, while those parents with low income mostly enroll their child or children into state-run monolingual schools (Anecdotal evidence from experience, author Syahabuddin 2010).

Acehnese people share the same problem with that of the majority of Indonesian parents regarding bilingual education: that is, most parents want their children to be taught bilingually but there are insufficient bilingual schools and teachers. In the context of bilingual education in Aceh, numbers of private schools have been occasionally providing bilingual programs for some decades. It is important to bear in mind that the levels of bilingual programs offered vary from one private school to another. Although these schools do not claim schools with bilingual programs, one can argue that the programs they provide are similar to Developmental Bilingual Education. Just after the 2004 tsunami, when Aceh received support and help with education from international agencies, some international-funded bilingual schools suggested the idea of bilingual education for the re-building after the 2004 tsunami. Starting from that time people have become familiar with the terms ‘bilingualism’ and ‘bilingual education’ and with what it means to enrol their children into one of these schools. One famous and favourite bilingual school in Aceh is Fatih Bilingual School, which has operated since December 26, 2006 (Forum, 2006). Apart from Fatih Bilingual schools in Aceh, there are state schools that provided bilingual programs. Those schools are called International Standard Schools (also known as Sekolah Berstandar Internasional, or RBI) and Candidate to International Standard Schools (Rintisan Sekolah Berstandar Internasional or RSBI). Those schools returned to normal standard schools (with no bilingual programs) when bilingual education was banned in Indonesia in 2013.

In general, the context of bilingual education in Indonesia and Aceh is argued to be different from elsewhere. In Aceh, the aim for students learning English within a bilingual school is not to prepare them for a mainstream class (for example, an English mainstream); rather, the aim is to prepare them to be able to function in two languages, that is, Bahasa Indonesia and English. This function is said to be needed for their future academic studies and jobs. It is suggested that the context of bilingual programs in Aceh tends to be what it is called ‘Bilingual education in majority languages’ (Baker, 2011), where a strong language or a national language of the region is used alongside English as the medium of instruction.

During the last two years since the re-building after the 2004 tsunami, people in Aceh have noted that students graduating from bilingual programs have outperformed students graduating from monolingual state-programs. (This is anecdotal evidence and it is neither supported nor refuted by any reputable studies in Aceh). A question arises: In Aceh, are students learning in bilingual programmes performing and achieving better than students learning in state-run monolingual programmes? Based on this question, the present study will conduct research on bilingual and monolingual students’ achievement in English writing, English comprehension, and in attitude towards and behaviour surrounding learning English; and this study will also investigate teachers’ attitudes and behaviour regarding their bilingual and monolingual students’ ability in English.

Bilingual and Monolingual Education

This study is concerned with the teaching of English in both bilingual and monolingual schools and so the background focuses mainly on English instruction. The published literature, however, is not so finely focused and so the present review includes some more general research relating to bilingual and monolingual instruction.

Bilingual education, in broad sense, is termed as the use of two languages as a means of teaching and learning instruction (Anderson & Boyer, in Romaine, 1995) of selected subjects (Margana, 2009). More specifically, bilingual education is defined as education that aims to promote bilingual competence by using both languages as the media of instruction with respect to three features: linguistic goals, pedagogical

approaches, and levels of schooling, for significant portions of the academic curriculum (Genesee, 1987; Met, 1998).

In Aceh, as in other provinces in Indonesia, bilingual education aims to promote the maximum acquisition of English due to changes in the Indonesian educational context, which were the result of the increased role of English as an international language and the language of technology (Margana, 2009). In practice, most bilingual classroom settings use English as the medium of instruction, whereas Indonesian is used as the medium of instruction in all monolingual classroom settings. The monolingual classroom setting in Aceh, on the other hands, involves classrooms where the medium of instruction to explain English words, sentences, stories and context is Bahasa Indonesia (Classroom Observation, done by Khairiah Syahabuddin, 17 January 2011).

There are a number of reviews of the literature relating to bilingual education (see for example Genesee, 1987; Met, 1998; Rossell & Baker, 1996). Rossell and Baker (1996) stated that “of 300 program evaluations read, only 72 (25%) were methodologically acceptable” (p.7). They also found that “on standardized achievement tests, transitional bilingual education [i.e. a bilingual education type which teaches students in their native languages as a preparation for main stream English-only classrooms] is better than regular classroom instruction in only 22%” of the programs. So the question of which program is better, bilingual or monolingual, is open to question, as it is in Banda Aceh. Another major problem with all these studies is that the tests were done with scores based on True Score Theory. The items are not ordered by difficulty to form a scale and the scores based on them are non-linear, and often not unidimensional either (see Wright, 1999). Michell (1990, 1999), a measurement expert, claims that with True Score Theory, one cannot even claim that a student with a higher percentage score has more achievement than a student with a lower percentage score, by logic. In True Score Theory, a typical standard error of measurement (SE) for percentage scores with a split-half reliability of 0.75 and a standard deviation of 20 is $SE = 20 \sqrt{1-0.75} = 10$. That is, a student with a score of 80% does not necessarily have a better achievement than a student with a score of say 70%, within the error of measurement.

A problem with many of the published long term studies (such as over one year or more) is making a true causal inference when so many extraneous variables are not controlled between pretest and posttest. This means that the true cause of any alleged improved achievement is probably a combination of variables, such as differences in time-on-task, differences in English homework, differences in English after-school-hours television watched and English books read, not just a difference between school bilingual instruction or school monolingual instruction. This leads to the proposed study in Banda Aceh and what is needed to be done. The proposed study needs to use the best available measurement procedure (namely Rasch measurement), and the best available Rasch computer program (Rasch Undimensional Measurement Model, RUMM) (Andrich, et al., 2010), and it needs to control for any extraneous variables that may affect the student outcomes in writing, reading comprehension and attitude/behaviour between pretest and posttest.

Rationale

Aceh is unique as one of 33 provinces in Indonesia, because it has been given privileges which have not been given to any other provinces in Indonesia (Government, 2000). Aceh province has been given three unique privileges in education, religion, and culture. This means that the province has freedom to manage its governance in terms of education, religion and culture in order to meet its needs. In line with the privileges in education, Aceh has added Islamic wisdoms as local content under Indonesia's curriculum for primary and secondary education. For tertiary education, Aceh has allocated a great amount of funding for education, such as scholarship provisions for Acehnese students to study abroad. In addition to that, at present, the Acehnese government welcomes any discourse and thoughts to enrich Acehnese education. This study will inform the Acehnese government and Acehnese people of any possible benefits of bilingual education in Aceh. Should the Aceh government agree on the idea of implementing a compulsory bilingual program into the education curriculum, it can pass on the information on bilingual education to state schools. It is not that the Indonesian government should base its policy on this one study but, if the present study shows a positive outcome, then further studies should be implemented to help develop education policy in Aceh.

While it is commonly believed by people in Banda Aceh that bilingual schools are better than monolingual schools, there is no clear evidence for this with Banda Aceh data. The students begin learning English in Year 7 and the common belief is that bilingually taught students are better at English writing, better at English reading comprehension, and have a better attitude and behaviour with regard to learning English than their counterparts in schools where teaching is monolingual. To answer their beliefs with evidence, this study compares bilinguals and monolinguals on Rasch-created linear measures of the three English outcomes over a two month period.

Statement of Purpose

The purpose of the present study is three-fold. The first is to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English text writing, English reading comprehension, and attitude and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual/monolingual). The second is to investigate the students' attitudes and behaviour in relation to the use of languages of instruction (combined English and Bahasa Indonesia vs. Bahasa Indonesia only) on their ability in English. The third is to investigate the students' perceptions about their learning English as a foreign language through their written comments.

Research Questions

1. Do first year middle state-school bilinguals in Banda Aceh achieve better in English Reading comprehension and English text writing than those of monolinguals?
2. Do first year middle state school bilingually-students at bilingual schools in Banda Aceh have better attitudes and behaviours with regard to learning English than those at monolingual schools?
3. What are the attitudes and behaviours of first-year middle state school bilinguals with regard to bilingual and monolingual education in Banda Aceh in terms of learning English?
4. What are first-year middle state school students' perceptions with regard to their learning of English as a foreign language?

Significance

The study is important for the following four reasons. One, there have been no previous studies comparing bilingual and monolingual educational outcomes with Aceh data or with any data in a Muslim province of Indonesia. Since the re-building of the education system in Aceh after the 2004 tsunami, the people of Aceh are asking the question about which type of instruction is better and the present study will thus provide data in relation to answering this question.

Two, the study has sufficient sample size and sufficient power in both the control group (N=386) and the experimental group (N=394) to give a definitive answer to the research questions. There are adequate controls and monitoring for extraneous variables such as time-on-task, content taught, English homework given, watching English television after school hours, reading English books after school hours, and after-school English learning. Consequently, the present study has the capacity to provide a comparison result that is credible.

Three, there have been no previous similar studies done with linear, unidimensional Rasch measures comparing bilingual and monolingual educational outcomes and all known previous studies in other countries comparing bilingual and monolingual educational outcomes have used non-linear scores based on True Score Theory measurement. The present study thus does not only provide good quality data to answer the research questions but it also uses current world's best measurement practice to produce linear, unidimensional measures of three educational outcomes (English writing, English reading comprehension, and attitudes and behaviour with regard to learning English).

Four, the present study should give direction for some future research into bilingual and monolingual education that should lead to good policy development in Aceh. The present study has clear educational policy implications. If the present study shows that there is a clear advantage in educational outcomes for bilingual or monolingual education, then this will 'drive' further research and provide important information for education decision-makers in Aceh. The present study can be regarded as an important future foundation for the educational system in Aceh on how to lead English teaching and learning in Aceh, especially in bilingual education.

Limitations

There are four main limitations to the present study. One, strictly, the data and the results only apply to Aceh province and not to any other Muslim provinces in Indonesia. Even though Aceh province has some similarities to other Muslim provinces in Indonesia, there are some substantial differences in other provinces, such as being more secular; these differences may make differences in educational outcomes possible. On the other hand, it might be assumed that the results applicable to Aceh province should be 'more-or-less' applicable to other Indonesian provinces, until proven otherwise.

Two, particularly, the results of the present study are not applicable to other Muslim countries because Aceh province has a unique cultural and trading background, and because of the redevelopment of the educational system following the 2004 tsunami. Other Muslim countries, especially in the middle-east, have long histories of Muslim economic development, education and culture that are different from Aceh and their educational systems are different. The differences might make for some differences in the way that bilingual and monolingual education occurs and hence for differences in educational outcomes.

Three, exclusively, the data for the present study only apply to Year 7 of middle school and not to any other years. The results will be generalizable to other Year 7 students in Aceh, Indonesia, but not to other schooling years in Aceh, Indonesia, because the tests were conducted for the Year 7 students only, not to other schooling years' students. It might, however, be expected that there will be little differences across other years. If the results are true for Year 7, then they are likely to be true for Years 8-12, but only research can tell.

The reason of choosing the Year 7 (semester 2) students for the study as the respondents was because they were beginners in learning English, and therefore, were assumed to share a similar starting point in learning English, and any experience gained during the study could then be associated with the medium of instruction. Under Curriculum 2006, English is taught as part of the curriculum in Year 7 semester 1. The present study investigated students' English ability and behaviours with regards to language of instruction. Year 7 (semester 2) students were considered suitable for the

study because they were beginners. They just had learned English in that schooling year (Year 7) and just had been learning English for 5 months prior to the study. They shared similar abilities in English and were considered to have a similar novice command in English. Within the two months of participating in the study, the two groups of students (Bilingual and Monolingual) were more likely to experience in-class English learning differently, due to the view that the language of instruction was different. Students learning English in a bilingual program were taught using combined English and Bahasa Indonesian. Students learning English in a Monolingual program were taught with the majority of Bahasa Indonesia and small amount of English. If other year students (for example, Year 8 students) were taken as the respondents of the study, they were not beginners anymore in learning English and the language of instruction might not be the only difference between them in regards to their English ability and behaviours because they would have had differing relative amounts of English knowledge and ability, and perhaps very different English learning behaviours. The two month difference between pretest and posttest may not have sufficient effect on them.

The present study used pretests and posttests separated by a two-month period and, while this should be enough time for differences in educational outcomes to occur, it may be that time differences over longer periods such as one or two years would produce larger differences. A problem for larger time-span research is that add-on variables such as after-school differences in time-on-task and after-school differences in exposure to English writing, speaking, reading, and watching television and movies may be just that or they may be due to the bilingual program. It is difficult to administer experiments to determine ‘what causes what’ here.

Four, solely, the data only apply to two English skills, namely, English Writing and English Reading Comprehension and not to the other two English skills, namely, English listening and English speaking. Further research is needed with a similar format and linear measures to determine the result for English listening and speaking.

Definition of Terms

The following definitions explain the meaning of the terms as used in the context of the present study.

Bilingualism

The term 'bilingualism' is reserved to describe two languages used by individual students, regardless of students' fluency in each of the languages.

Bilingual Education

Bilingual education is defined as education that aims to promote bilingual competence by using both languages as media of instruction with respect to three features: linguistic goals, pedagogical approaches, and levels of schooling, for significant portions of the academic curriculum. More specifically, for the purpose of the present study, it is defined as the teaching of English by using a combined English and Bahasa Indonesia with equal percentage of time.

Bilingual Classroom Setting

A bilingual classroom setting is an English language classroom in which English is used as the medium of instruction, whereas Indonesian is used as the medium of instruction in all other classes.

Monolingual Education

A monolingual classroom setting is an English language classroom in which Bahasa Indonesia is used as the medium of instruction to explain English words, sentences, stories and context. More specifically, for the purpose of the present study, it is defined as the teaching of English by using a majority of Bahasa Indonesia and a small amount of English by time.

English Reading Comprehension

English Reading Comprehension is the ability of second language readers to obtain meaning from texts by actively using both lower and higher skills to decode the smaller elements and construct the meaning. By using their schemata (previous or background knowledge), they are able to understand the main idea, sequence the order, and obtain detailed information. The English reading comprehension in this study represents student English achievement and it is based on the Bloom Taxonomy.

English Reading Comprehension Test

The English Reading Comprehension Test used in the present study consists of 12 multiple choice items and three written items to be answered by the students after reading a given piece of English text (see the full test in Chapter Three). A special scoring rubric in which scoring is ordered by item difficulty for the multiple choice items and ordered by quality for the written answers was designed for the test (see the scoring rubric in Chapter Three). This scoring is consistent with Rasch measurement principles and is used to create a linear, unidimensional measure for the variable English Reading Comprehension.

English Text Writing

English Text Writing is the ability of second language writers to produce written texts by actively using writing strategies, techniques and skills, which have been acquired during their English language learning instruction, whether it be in bilingual or monolingual classes. The English text writing in this study represents student English achievement and it is based on Bloom's Taxonomy (Bloom et al., 1956).

English Text Writing Test

The English Paragraph Writing Test consists of two compulsory topics on which the students are asked to write several paragraphs in English (see the full test in Chapter Three). A special scoring rubric in which scoring on three aspects of writing is ordered by quality was designed for this test (see Chapter Three). This scoring is in line with

Rasch measurement principles and is used to create a linear unidimensional measure of English Writing Quality.

Attitude and Behaviour with regard to Learning English

The Attitude and Behaviour with regard to Learning English Questionnaire (see Appendix E) contains 18 items: three on tasks for listening, three on tasks for speaking, three on tasks for reading, three on tasks for writing, three on tasks for student/student relationships, and three on tasks for student/teacher relationships. Each item was answered from the two perspectives of attitude and behaviour, that is ‘ideally, this is what I think should happen’ (attitude and easier); and ‘this is what actually happened’ (behaviour and harder). The full questionnaire is given in Chapter Three. Response category ‘most or all of the time’ was scored 3, response category ‘some of the time’ was scored 2, and response category ‘never or rarely’ was scored 1. This scoring is ordered in line with Rasch measurement principles and was used to create a linear unidimensional measure.

Middle School

Middle school in Aceh is the level after secondary school. Middle school students are 12-15 years old. The duration of middle school is three years, and English classes are begun in the first semester of the first year of this level.

True Score Theory Measurement

True Score Theory is a way of measuring variables in the social sciences (and education) which claims that the observed total score obtained by a person on a set of test or questionnaire items is made up of a ‘true score’ and a random error score. The scale created by True Score Theory does not contain equal units of measures and is therefore non-linear. That is, the difference between, for example, 50% and 60% does not represent the same amount of variable difference as between 70% and 80%. True Score Theory scores are commonly considered to have at least six problems: (1) non-linearity; (2) multi-dimensional with ‘noise’; (3) item difficulties not ordered; (4) person ‘measures’ and item difficulties not ordered on the same scale; (5) the ‘measures’ are test (item content) dependent; and (6) the ‘measures’ from different tests, even on the

same topic, cannot be validly added or linked onto a single scale (see Michel, 1990, 1999; Smith, 1996; Waugh & Chapman, 2005) .

Rasch Measurement

In Rasch measurement, items are ordered from easy to hard on a continuum and their difficulties are calculated on a linear scale (a log odds scale). The person measures are calculated on the same linear scale. An important point to understand is that when the data fit a Rasch measurement model, the differences between the person measures and the item difficulties can be calibrated together in such a way that they are freed from the distributional properties of the incidental parameter, because of the mathematics involved in the measurement model. This means that ‘scale-free’ measures and ‘sample-free’ item difficulties can be estimated with the creation of a mathematically objective linear scale with standard units. The standard units are called logits (the log odds of successfully answering the items) (This has been taken from Waugh, 2003, 2005, 2010a; Waugh, 2010b).

Unidimensionality of Variables

In the present study, unidimensional measures are created for English Writing, English reading Comprehension, and Attitude and Behaviour about Learning English. These measures involve a variety of aspects including low order thinking (such as knowing facts and basic comprehension), higher order thinking (such as analysis, synthesis and evaluation), low and high order attitudes, and physical dexterity and, in this sense, cannot be unidimensional. With Rasch measurement, unidimensional means that a single parameter for each person (person measure) can be created as applying to all of the scale items, that a single parameter can be created for each item (item difficulty) applying to all the persons measured on the same scale, and that these parameters can be applied to accurately predict each person’s response to each item.

Person Separation Index

Person Separation Index is an index ranging from 0 to 1 that shows the proportion of observed variance considered to be true. A high value of the index indicates that measures of the respondent’s ability or preference are sufficiently well

separated along the scale in relation to the errors of measurement. It is “structured as the ratio of estimated observed variance among persons, using estimates of their locations (measures) and the standard errors of these locations (measures)” (Andrich & van-Schoubroeck, 1989, p. 483) . The Person Separation Index is interpreted like Cronbach’s alpha which measures the internal reliability of non-linear scales (Cronbach, 1951).

Item Thresholds

Item thresholds show the location on a continuum whereby it is likely a person will obtain a particular score. More specifically, thresholds are points between adjacent response categories where the odds of answering in either category are 1:1. With three response categories there are two thresholds and with four response categories there are three thresholds. Thresholds should be ordered in line with the ordering of the response categories showing that the responses are answered consistently and logically (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Standardised Residual

Residuals are differences between the expected response according to the Rasch measurement model and the actual response. The standardized residual is the residual divided by its standard deviation. When there are many standardized residuals, then the mean should be close to zero and the standard deviation should be close to one, when the data fit the Rasch measurement model (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Response Category Curves

Response Category Curves show whether items have been answered logically and consistently. The actual curve that is produced shows the relationship between the probabilities of answering each category in relation to the specific measure. For example, the ideal curve for an item with three response categories shows that when the measure is low, the probability is high that the participant response is low (category one). As the measure increases, the probability of answering category one decreases and the probability of answering category two increases. As the measure increases further

still, the probability of answering category two decreases and the probability of answering category three increases (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Differential Item Functioning

Differential Item Functioning refers to items that give different success rates for two or more groups, at the same ability level (Holland & Wainer, 1993). Masters (1988a) states that item bias occurs if an item's estimated difficulty is significantly greater when calibrated on one sub-group than when calibrated on the other, resulting in the item being considered 'biased' with respect to those two sub-groups. In other words, test bias can occur when the test requires different information or knowledge than that being tested, causing test scores to be less valid for a particular group of test-takers (see also Penfield & Lam, 2000).

Item Characteristics Curves

Item Characteristic Curves show how well the items differentiate between persons with differing measures. An ogive curve (see Figure 1.1 for an ogive curve) is produced for each item showing the relationship between the expected response score and the particular measure (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

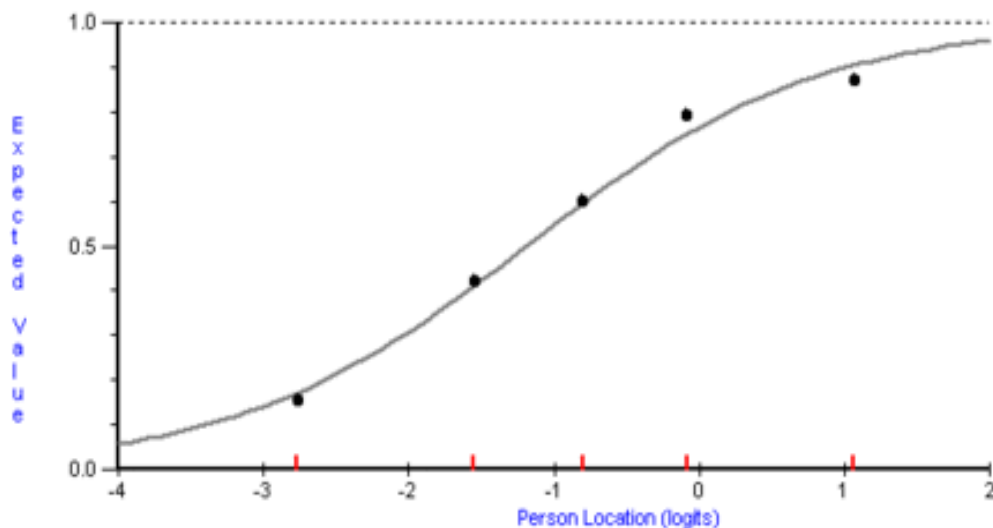


Figure 1.3 An Ogive Curve

Structure of the Thesis

This thesis is presented in twelve chapters. Brief summaries of the contents of Chapters Two to Twelve are given below.

Chapter Two is the literature review. The chapter provides a discussion on bilingualism and bilingual education in the world. It includes reviews of research results relating to bilingual and monolingual education and also the results of individual studies in western and Muslim countries. The research literature is critically analysed to show the gaps in the research and its deficiencies in implementation of research projects.

Chapter Three presents the conceptual framework of bilingual and monolingual teaching and it explains the expected outcomes according to modern theory. A discussion on the deficiencies of measurement using True Score Theory is provided and Rasch measurement is explained as a strong improvement in line with world's current best practice. The structure of the measurement instruments for the three variables to be used in the present study, that is, the English Writing Test, the English Reading Comprehension Test, and the Attitude and Behaviour Questionnaire about Learning English, are described, along with the scoring rubrics for each variable.

Chapter Four presents the mixed-method design. This involves the quasi-experimental approach of intact control and experimental groups with pretest and posttest measures, and their associated samples and piloting of questionnaires. The design involves Focus Groups and video recording of several bilingual and monolingual classrooms, with their associated samples and piloting. The discussion expands on the control of extraneous variables in the quasi-experiment, test data collection, data entry and data analysis.

Chapter Five (Data Analysis, Part 1) presents the Rasch data analysis results for English Reading Comprehension. The results of the RUMM2030 computer program output involving tabular data such as overall item and person fit, individual item and person fit, item-trait interaction (dimensionality), thresholds, and reliability (Person Separation Index) are presented and explained. Graphical output from the RUMM2030 program involving Response Category Curves, Item Characteristic Curves (ogives) and targeting graphs are also presented and explained. Similar RUMM2030 output results

for English Writing are presented in Chapter Seven (Data Analysis, Part 3) and for Attitude and Behaviour to Learning English in Chapter Nine (Data Analysis, Part 5).

Chapter Six (Data Analysis, Part 2) presents the ANCOVA and ANOVA results for the quasi-experiment interaction and main effects for English Reading Comprehension. Similar output results for English Writing are given in Chapter Eight (Data Analysis, Part 4), and output results for the Attitude/Behaviour Questionnaire are presented in Chapter Ten (Data Analysis, Part 6).

Chapter Eleven (Data Analysis, Part 7) presents the data analysis of the students' written comments.

Chapter Twelve presents the discussion and implications of the results of this study in the light of current literature findings. The chapter concludes with implications for bilingual education in Aceh province for teachers, students, policy makers and future research.

The next chapter reviews literature on bilingualism and bilingual education.

CHAPTER TWO

BILINGUALISM and BILINGUAL EDUCATION

There is a great deal of literature covering many years past on bilingualism and bilingual education in the relevant journals, books and reports. Clearly, bilingualism and bilingual education has been, and continues to be, an important and controversial topic in many countries across the world. There are too many studies and reports to summarise all of them for the present chapter of this study. This chapter briefly revisits the nature of bilingualism, the nature of bilingual education, and the nature of bilingual education in Indonesia and summarises some of the more relevant and recent research (2003-2013) on these topics from countries across the world. In this chapter, there is a selection and summary of research commentary on the nature of bilingualism, the benefits of bilingualism, the disadvantages of bilingualism, the nature of bilingual education, models of bilingual education, bilingual education in Indonesia and the recent 2013 judicial banning of bilingual teaching in Indonesia.

The Nature of Bilingualism

What does it mean to be a bilingual?

To define the term bilingual is not easy. It has been defined as one's knowledge and one's use of two languages (Butler, 2013). The term might mean that one's ability in the two languages is balanced. However, this is rarely the case, due to the fact that the level of one's bilingualism differs from one bilingual person to another. In regards to the four language skills (listening, speaking, reading, and writing), for example, a bilingual person might be strong at the two productive skills (speaking and writing), but might be weak at the other two receptive skills (listening and reading) (Baker, 2011), or *vice versa*. So, a question arises, "How bilingual is bilingual?" In line with this, Shih (2012) believed that 'a balanced bilingual' is a myth and referred to it in the form of circles by analogy (see Figure 2.1).

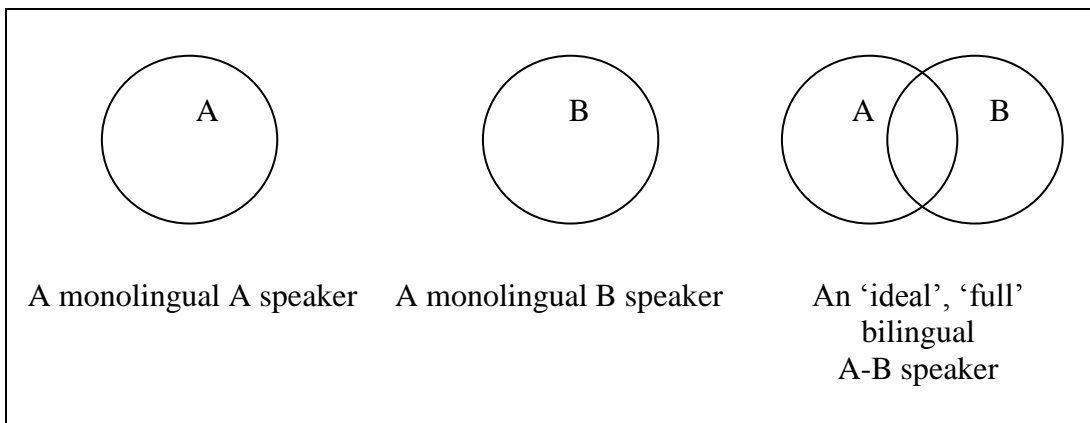


Figure 2.1 *A Mythical View of the 'Ideal' or 'Full' Bilingual (Source: Shih, 2012, p.4)*

The size of the circles in Figure 2.1 indicates, by analogy, that the level of proficiency in each language is the same for a bilingual speaker and for his/her bilingual ability. In reality, one's bilingual ability is rarely 'balanced'. Although a balanced bilingualism is an ideal concept, only a few bilingual speakers can attain it (Baker, 2011; Wei, 2000). More often, one's bilingual ability varies from another bilingual and so the circle analogy should be more like those depicted in Figure 2.2.

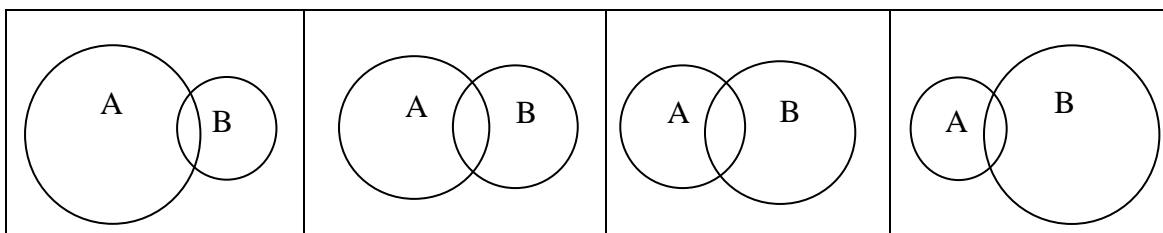


Figure 2.2 *More Realistic Conceptions of Bilingualism (Source: Shih, 2012 p. 5)*

Grech and McLeod (2012) have summarised the breadth of definitions of bilingualism into three categories:

1. Bilingual exposure from birth;
2. Using more than one language in day-to-day functioning; and
3. A continuum of use and proficiency in more than one language.

Due to the difficulty of representing a simple definition of bilingualism, Baker (2006, pp. 3-4) outlined eight dimensions of bilingualism. These dimensions are summarised and explained here due to their contribution towards the question posed earlier, 'How bilingual is bilingual'?

Eight Dimensions of Bilingualism summarised from Baker (2006, pp. 3-4):

1. Ability

Ability here refers to whether bilinguals use both languages actively or passively. Some have the ability to speak and write actively in both languages. This ability is called productive competence. Others, on the other hand, have the ability to listen and read passively in both languages. This ability is called receptive ability. Some bilinguals might only receive understanding from hearing and reading but they are not ready to produce the form of utterances (speaking ability) and written discourses (writing ability) for the two languages. Valdes et al. (2003), quoted in Baker (2006), affirm that ability is on a continuum. Bilinguals vary their ability with dominance of one language over the other, and vary in their development of one language over the other.

2. Use

The places and settings where the languages are obtained vary. Some bilinguals might acquire an additional language at home, school, on the street, and/or phone, Facebook, or Twitter. The languages are often used for different reasons or purposes.

3. Balance of two languages

Most often the use of the two languages is imbalanced, and one language is dominant over another.

4. Age

Bilinguals may acquire languages at different ages. Some acquire the two languages from birth. This is called infant or simultaneous bilingualism. Some acquire one language after the age of three. This is called sequential or consecutive bilingualism.

5. Development

The development of the languages varies. When a language is fully acquired, while the other has just started, it is called incipient. When the second language is

developing, the bilinguals are called ascendant. When one of the languages is decreasing, the bilinguals are called recessive bilinguals. When a language starts to stop, the bilinguals are called language attrition bilinguals.

6. Culture

If the person learning a language is also learning the culture of the language, the process is called acculturation. Bicultural competence tends to relate the knowledge of the two cultures of the two languages as well as the feelings, attitudes, and behaviours of the bilinguals.

7. Context

The places where the bilinguals live vary. Some bilinguals live in a bilingual/multilingual place where two or more languages are used daily (known as endogenous context). Some live in places where there are no second language communities (known as exogenous contexts). Some bilinguals have contact with second language communities through vacations, phone, email, Facebook, Twitter, television, films, books and so on (known as network contexts). Some contexts may be subtractive, where the home language is replaced by the community language due to the politics of the majority language users. Other contexts are additive, where a person adds a new language without any cost occurring; this is the case for some elite or prestigious bilinguals.

8. Characteristics

Elective bilingualism is a characteristic of people who prefer to learn a second language in a classroom. They add a second language without losing their first language. Circumstantial bilinguals learn another language to function effectively due to their life demands. For this characteristic, there is a concern at losing their first language, – a subtractive context. These two characteristics (elective and subtractive) are important, because they lead to practice, to status, to politics, and to power issues among bilinguals.

In line with the eight dimensions, Butler (2013) provided a concise table which is called ‘Classical Typology of Bilingualism’, and it is set out below (see Table 2.1).

Table 2.1
Typology of Bilingualism Classical

<i>Point of Focus (Dimension)</i>	<i>Typology</i>	<i>Definitions</i>	<i>Characteristics of Second Language Acquisition</i>	<i>Related issues and educational implications</i>	<i>Additional complications in multilingual contexts</i>
Relationship between proficiencies in two languages	Balanced; Dominant (Peal & Lambert, 1962)	Achieving equal level of proficiency in L2 and L1 (balanced); L2 proficiency varies but not the same as L1 (dominant)	Functional differences; related to age factor	Conceptualizing and assessing language proficiency; Cummins’ threshold hypothesis and interdependent hypothesis; Semilingualism	Greater complexity in conceptualizing and measuring multilingual competences
Functional ability	Receptive; Productive	Understand but not produce L2 either in oral and/or written domains (receptive); understand and produce L2 (productive)	Functional and motivational differences	Language use irrespective of proficiency levels and identity	Greater diversity in functional differences across domains and across languages
Age of acquisition	Early; Simultaneous; Sequential; Late (Genesee et al. 1978)	Exposed to two languages from birth (simultaneous); Exposed to L2 after L1 has some foundation (sequential); became bilinguals during adulthood (late)	Maturational difference; schooling differences	Neurolinguistic differences; critical period hypothesis	Greater diversity in the acquisition order, can have multiple L1s and/or L2s
Organisation of linguistic codes and meaning unit(s)	Compound; Coordinate; Subordinate (Weinreich, 1953)	Two sets of linguistic codes stored in one meaning unit (compound); stored separately (coordinate);	Functional differences; differences in form-meaning mapping	Difficulties with operationalising distinctions and testing differences	Greater complexity and diversity in multilingual memory organization according to typological differences

		L2 is assessed through L1 (subordinate)			among languages and proficiency levels
Language status and learning environment; literacy support of L1	Elite; Folk (Fishman, 1977); Circumstantial; Elective (Valdes and Figueroa, 1994)	No or little additive value of L1 as a language minority status (folk / circumstantial); additive value of L2 (elite/elective)	Differences in language status and value of bilingualism	Support for literacy in L1 and L2 literacy development	Greater diversity in social values attached to multiple languages
Effects of L2 learning on the retention of L1	Additive; Subtractive (Lambert, 1974)	L2 as enrichment without loss of L1 (additive); L1 is replaced by L2 (subtractive)	L2 as enrichment with or without loss of L1; status of a language in a given context	Social value of L1 greatly influences the retention of L1; support for literacy in L1 and L2 literacy development	Greater complexity of learning an additional language from previously acquired languages; greater diversity of status across languages
Cultural identity	Bicultural; L1 monocultural; L2 accultural; Deculturated (Hamers and Blanc, 2000)	Cultural identity shaped by two cultures (bicultural); identity in one culture (L1 monocultural); loss of L1 culture (L2 accultural); identity in neither cultures (deculturated)	Differences in acculturation process	High bilingual competence does not necessary coincide with dual identity	Multiple cultural identities coexist irrespective of competences

Source: (Shih, 2012, pp. 113-114)

This table gives us some more comprehensive ideas on what it means to be a bilingual, as there are numbers to consider, such as, point of focus (dimension), typology, definitions, characteristics of second language acquisition, related issues and educational implications, and additional complications in multilingual contexts.

To sum up on the terminology of bilingualism, the literature suggests that perhaps a better way to label a bilingual is to consider the dimensions: ability, use, balance of the two languages, age, development, culture, context, and characteristics.

What are the Benefits of Being a Bilingual?

It has been documented through a great number of research studies that people who are bilingual and/or have received bilingual education outperform people who are monolingual and/or have received monolingual education on cognitive matters.

It has been documented through a great number of research studies that bilingualism and bilingual education outperforms monolingualism and monolingual education on cognitive matters. Among them is executive control, metalinguistic awareness, working memory, metacognitive awareness, and divergent thinking. Cognition has been defined as the ability to act in line with goals on learning and the ability to resist inference Morton and Harper (2007). An increasing body of studies finds many positive effects of bilingualism on cognitive advantage and evidence has been found that bilingualism enhances the cognitive development of children in a variety of circumstances and cultures (see Bialystok, Craik, & Freedman, 2007); Diamond, 2002; Kave, Eyal, Shorek, & Cohen-Manfield, 2008).

Executive control

Particular evidence regarding cognitive development is on executive control functions. Executive control, which is believed to be a most important aspect of cognitive development (see Diamond, 2002), refers to a set of interrelated processes in the frontal lobes (Bialystok & Viswanathan, 2009). It consists of three components, namely, (1) shifting between tasks on mental sets; (2) updating and monitoring memory presentations; and (3) inhibition of dominant or pre-potent responses (Miyake, Friedman, Emerson, & Witzki, 2000). Cognitive development is needed so that the two languages [became] integrated with the linguistic circuits used for language processing, creating a more diffuse, more bilateral, and more efficient network that supports high levels of performance” (Bialystok, 2011, p. 233).

The effects of enhanced executive control develop earlier in bilingual children than in comparable monolinguals (Adi-Japha, Berberich-Artzi, & Lidnawi, 2010; Ellen

Bialystok, 2010; Carlson & Meltzoff, 2008) and have been found at all stages of the life span, beginning from infancy (see Bialystok, Craik, Klein, & Viswanahan, 2004; Carlson & Meltzoff, 2008; Costa, Hernandez, & Sebastian-Galles, 2008; Kovacs & Mehler, 2009; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011). Evidence for executive control comes from behavioural and imaging studies. Results of the behavioural studies indicate that interference occurs to participants' other language (Francis, 1999; Kroll & deGroot, 1997). Results of the imaging studies indicate that interference from the non-target language is important (Colome, 2001; Marian, Spivey, & Hirsh, 2003; Rodriguez-Fornells, Rotte, Heinze, Nosselt, & Munte, 2002; Wu & Thierry, 2010). These behavioural and imaging studies reveal that both languages of bilinguals are constantly active to some degree (Bialystok, 2011), and available when either of them is in use (Kroll, Bobb, & Wodniecka, 2006; Marian, et al., 2003; Rodriguez-Fornells, Balaguer, & Munte, 2006; Sumiya & Healy, 2004), even in strongly monolingual contexts. The brain controls any inhibiting attention and hinders intervention from one language to another in joint activation for two languages (Greenberg, Bellana, & Bialystok, 2013).

Metalinguistic awareness

Metalinguistic awareness has been defined as the ability to think about language in terms of its form and structure and the relation between the two to produce comprehensible meanings (Adesope, Lavin, Thompson, & Ungerleider, 2010). It can also be recognised as the ability to direct attention to the systematic characteristics of language, and an ability to mirror them (Lazaruk, 2007).

Research on metalinguistic awareness has been conducted by numbers of reseachers (Baker, 2006; Bialystok & Barac, 2012; Chow, McBride, Cheung, & Chow, 2008; Kieffer & Lesaux, 2012; Kirby, Desrochers, Roth, & Sandy, 2008; Laurent & Martinot, 2009; Lazaruk, 2007; Rauch, Naumann, & Jude, 2011; Rispen, McBride-Chang, & Reitsma, 2008). Studies on metalinguistic awareness have concentrated on three awarenesses: words awareness: phonological awareness (also known as sound system awareness) and syntactic awareness (also known as word-order awareness) (Chow, et al., 2008).

Research on word awareness has consistently shown that bilingual children retain an astonishingly good grasp of relations between words and languages and the meanings of the words in the two languages. An example is the study by Bialystok and Barac (2012). The aim of their study was to ascertain the nature of the experiences of bilingual learners that make a contribution to performance on metalinguistic awareness. Their results uniformly demonstrated that bilinguals' metalinguistic performance improved as the knowledge of both languages increased. This is in line with another study conducted by Laurent and Martinot (2009) on phonological awareness in children exposed to the early learning of a second language. It aimed to investigate whether this exposure to the second language highlights the learners' development of phonological awareness. They showed that bilingual learners show more significant positive results in their phonological awareness than do their monolingual counterparts (see also Laurent & Martinot, 2009). Regarding this ability, the study by Baker (2006) on metalinguistic awareness among bilingual learners supported the view that bilingual learners 'gratify' their heightened metalinguistic abilities, instead of only possessing a universal metalinguistic dominance over monolingual learners.

Working memory

Working memory is defined as a temporary storage of information to be used for performance of various types of tasks (Baddeley, 2003). It is also seen as memory at work, not just memory (Conway, Jarrold, Kane, Miyake, & Towse, 2007). A reasonable amount of research has been conducted on working memory (see Abu Rabia & Siegel, 2011; Biedron & Szczepaniak, 2012; Morales, Calvo, & Bialystok, 2013; Ransdell, Barbier, & Niit, 2006; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011). A study done by Biedron and Szczepaniak (2012) investigated the role of short-term memory and working memory in accomplished multilinguals. The analysis revealed that short-term memory and working memory abilities in the multilinguals were higher than in mainstream students. Just recently, Morales, et al. (2013) explored two studies comparing the performance of bilingual and monolingual learners on tasks requiring different levels of working memory. Both studies show that bilingual children received advantages over their counterparts regarding working memory.

Metacognitive awareness

Metacognitive awareness refers to the awareness a person has of their own knowledge with regard to their mental activities and learning strategies (Adesope, et al., 2010). The process of learning two languages, which includes two systems of vocabulary, syntax, phonology, and morphology, and the process of using them in meaningful and appropriate contexts, may tell the bilingual learners how they manage to use their two cognitive processes and learning strategies (Kemp, 2007). Over the past decade, research has been conducted on the effects of bilingualism on learners' metacognitive awareness. The majority has found that bilingual learners show greater cognitive awareness than the monolinguals do (see Judge, 2011; Ransdell, et al., 2006; Vorstman, Swart, Ceginskas, & Bergh, 2009).

Divergent thinking

Divergent thinking can be understood as one's mental process and methods which are used to trigger creativity with the help of various alternative results (Kharkhurin, 2008). Research that investigated the outcomes of bilingualism on a learner's divergent thinking shows that bilingual learners develop greater ability to trigger and process numerous unconnected notions (Ghonsooly & Shawqi, 2012; Kharkhurin, 2008, 2009).

Is there any Disadvantage to being Bilingual?

There have been concerns over possible disadvantages of bilingualism, especially in the early years of the practice of bilingualism in the USA. Early bilingualism was seen as dangerous, leading to language disorder and language delay. These concerns have been associated with a deficit viewpoint. This deficit viewpoint was based on the idea that a child's mind could only hold one language at the same time, that is, more than one language was too much to bear. That concern might have taken root from a quotation from a professor that is now seen as an historical and hysterical comment (Baker, 2011): "If it were possible for a child to live in two languages; at once equally well, so much the worse. His intellectual and spiritual growth would not thereby be doubled, but halved. Unity of mind and character would have great difficulty in asserting itself in such circumstances" (in Baker, 2011, p. 139). That

fear was believed by certain people who associated it with deficiencies in the size of vocabulary gained in the two languages; competence and performance of the two languages and language creations, meaning and imagery (Wei, 2010). However, this fear has proved to be incorrect (Baker, 2011). It is widely believed at the present that having more than one language does not bring any harm to a child's mental ability. In fact, much research has shown that bilingual children excel over monolingual children in several aspects of cognition, as already stated.

Added to the deficit viewpoint are another four fears that have been outlined by Baetens-Beardsmore (2003): parental fear, cultural fear, educational fear, and politico-ideological fear. Parental fear has been associated with the unreadiness of parents to allow their children to encounter bilingualism. This unreadiness could be due to their own uncomfortable experience of being an adult trying to learn a second or foreign language. This fear has also been associated with misguided information received by parents from educational 'experts' who seem to believe in the deficit viewpoint. This parental fear could hinder their children's encounter with bilingualism.

Cultural fear refers to any possibility of an unbalanced acculturation. There have been some concerns that bilingualism could harm students' language heritage and home country cultures. This concern was commonly raised by immigrant parents who came to live in a new country. Some parents worried that their children would adopt a new language together with the new culture and, as consequence, might neglect their heritage culture and heritage language in favour of the new culture or they would lose their language heritage and the culture of their ancestors. The community's apprehension has been argued that it was related to identity (Edwards, 2010): the society want the members of their community to excel in their schools and at the same time they yearn for their language heritage and cultures to be preserved. Their fears were due to the view that the immigrants were expected to replace home country characteristics with the new country's characteristics, in order to make them more like the new country citizens (Dicker, 2003). Regarding this, some schools in these new countries have provided several different bilingual schools that accommodate the needs of the children. It has been suggested that a well-balanced bilingual educational program could minimize this fear (Baker, 2011). The schools ensure that children can

participate well in the classroom, but can also preserve their local language and heritage, at least to some extent.

Educational fear is usually expressed by people who have conceived misguided assumptions about bilingualism and bilingual learning. The vast majority of them represent people who are not directly involved in bilingual research or practices. This expressed fear has less to do with education than with perception about national unity. This fear is concerned more with seeing bilingual education programs as a threat to the state of the majority language (Baetens-Beardsmore, 2003).

The Nature of Bilingual Education

What is Bilingual Education?

Perhaps, before we look into the definition of a bilingual education, it is beneficial to consider what we know about bilingual education and mother tongue development, as has been raised by Cummins (2003), together with some answers, on the following points.

1. Bilingualism has positive effects on children's linguistic and educational development.
2. The level of development of children's mother tongue is a strong predictor of their second language development.
3. Mother tongue promotion in the school helps develop not only the mother tongue but also the children's abilities in majority school languages.
4. Spending instructional time through a minority language in the school does not hurt children's academic development in the majority school language.
5. Children's mother tongues are fragile and easily lost in the early years of school.
6. Negotiation of identity is a crucial factor in minority children's academic success (pp. 61-64).

With the above baseline in bilingual education and mother tongue development, bilingual education has been seen as a controversial and misunderstood matter (Freeman, 2007). Cazden and Snow (1990, p. 9) put it as "a simple label for a complex

phenomenon” and Butler (2013, p. 110) adds “a highly complex social, psychological, and linguistic phenomena”. Its controversial nature is due to the confusion regarding the same term being used for several different programs. The programs could vary in terms of different ideological orientations towards linguistic and cultural diversity, different target populations, and different goals for those target populations (Hornberger, 1991).

In broad terms, bilingual education has been defined as the use of two languages (native and additional language) as the medium of instruction (Brisk, 2005), with the ultimate goal to achieve additive bilingualism, biliteracy and cultural diversity, but the form of the bilingual education can vary greatly (Cummins, 2000). To determine which type of bilingual education is used depends on various variables, and the names of a few are: the native language of the students, which language to take, the language of instruction, and the linguistic goals for the program.

What are the Existing Bilingual Education Models and Practices?

Bilingual education has a diversity of aims, practices and contexts, and where bilingual education has been applied in one country, it has different aims and context to another. For example, bilingual education in the USA or Canada is different from a bilingual context in Malaysia or Indonesia. Bilingual education in the USA or Canada has been associated with minority language students who learn curriculum content in one majority language (for example, English) and in one minority language (for example, Spanish) in the USA context; and for two majority languages, namely French and English in the Canadian context. Bilingual education in Malaysia or Indonesia has been associated with majority language students learning curriculum content in two languages, where one language is a strong ethnic language (namely Bahasa Malay, in the Malaysian context) or a national language (namely Bahasa Indonesia, in the Indonesian context) and the other is English as a second language (in the Malaysian context) or foreign language (in the Indonesian context).

Due to the different aims of bilingual education in different countries and contexts, the perspectives of the different socio-historical contexts need to be considered (Lin & Man, 2009). Ferguson, Houghton, and Wells (1977) provided ten different goals of bilingual education:

1. To assimilate individuals or groups into the mainstream of society; to socialize people for full participation in the community;
2. To unify a multilingual society; to bring unity to multi-ethnic, multi-tribal, or multi-national linguistically diverse states;
3. To enable people to communicate with the outside world;
4. To provide language skills which are marketable, aiding employment and status;
5. To preserve ethnic and religious identity;
6. To reconcile and mediate between different linguistic and political communities;
7. To spread the use of a colonial language, socializing an entire population to a colonial existence;
8. To strengthen elite groups and preserve their privileged position in society;
9. To give equal status in law to languages of unequal status in daily life; and
10. To deepen an understanding of language and culture.

Based on the above ten aims of bilingual education (Baker, 1996) classified bilingual education into four categories:

1. Typical language(s) used by the child in daily life;
2. Typical language(s) in the classroom;
3. The educational and societal aim(s) of the program; and
4. The probable outcomes of the program.

Based on the above ten categories, bilingual education can be classified into three general types and then re-categorised into eight models or programs that have been in practice throughout the world (based on Baker, 2006, 2011). The three general types are: (1) Maintenance programs, (2) Transitional programs, and (3) Enrichment programs.

The specific ten models or programs (summarised from Baker, 2011, pp. 211-250; and (Ovando & Combs, 2012, pp. 35-44) are:

1. Mainstreaming/Submersion education (also known as ‘sink or swim’ method). By ‘submersion’, it metaphorically means that a child is allowed to swim in

the sea of knowledge, without any help. It is believed that the child will quickly learn how to swim in order to survive. In classroom situations, children will try to learn a totally new language without help, and it is hoped that they can cope with that in order to 'survive' in a language classroom. Submersion education is a label used to describe education for a language minority (in the USA context) who are placed in mainstream education with the majority language students. The language of the classroom is the majority language, while that of minority language students is other than English (for example, Spanish or French). Both teachers and students use a majority language (such as English). Thus, the minority language students whose first language is not English are forced to learn English without any help prior to the admission to the classroom. These students need to 'swim' to survive. This type of bilingual education has been practised in countries like the USA.

2. Submersion with withdrawal or pull-out, or sheltered English, or content-based ESL classes. In submersion with withdrawal/pull-out, the language used is the majority language (such as English). The minority language students are assigned to learn the majority language (English) with help under a second language curriculum. These lessons are taught separately from the submersion class. The submersion teacher teaches the academic content and the teacher with the English as a Second Language qualification teaches English. The students are withdrawn or 'pulled-out' from their mainstream class every day for an assigned time for a couple of years. In the submersion with sheltered English, the program contains only minority language students and is conducted in a majority language (in English for non-native students). The teacher uses a simplified form of the majority language and accepts contributions from the students in their native language but does not support the native language of the minority students. In the submersion with content-based English as a Second Language, the teacher teaches both the academic content in the majority language, and the majority language as a second language at the same time. The teachers have both academic content and English as a Second Language qualifications. This type of bilingual education has been practised in countries like the USA.

3. Segregationist education. This education is for minority language students only and is conducted in the minority language, but it is uncommon. This type occurs where minority language speakers are denied access to those programs or schools attended by

majority language students. Such education can be through law and practice. Segregationist bilingual education is aiming at prohibiting the minority language students acquiring the majority language in order that the minority language people do not interfere with the majority language elites. This type of bilingual education has been practised in countries such as South Africa, prior to Nelson Mandela becoming President.

4. Transitional bilingual education (early exit or late exit). This type is also known as Mainstream Bilingual Education or Developmental Bilingual Education. As the term suggests, this type of bilingual education prepares the minority language students with the majority language before they are ready for the mainstream classes. Containing the minority language students only, the students are temporarily allowed to learn academic content in their native language for a maximum of two years. After this period of time, the students join the mainstream classes where the language used is the language of the majority without any language of the minority. For the late exit transitional bilingual education, the student's native language is used in teaching academic content for longer than two years. It places less emphasis on exiting language students from the bilingual program as soon as possible. Students in class receive content area instruction in both languages (40% majority language and 60% native language) until grade 6. Teachers or their assistants need to be bilingual. The aim of this program is majority language monolingualism. This type of bilingual education has been practised in countries like the USA.

5. Mainstream education (with a foreign language taught as a subject). The mainstream education is for students learning academic content using their own home or native language as a majority language and some other foreign language is also learnt. For this class, a foreign language is regarded as a subject (as Maths, Social Sciences, and the like) and the foreign language learnt is not used as a medium of instruction. This type of bilingual education has been practised in Canada, with programs called 'core French' and 'drip-feed'.

6. Separatist education. Separatist education promotes monolingualism in the minority language as well as promoting monoculturalism where possible. This is a way to protect a minority language from being overrun by the language majority for

political, religious, or cultural reasons. This type of bilingual education has been practised in small isolationist religious schools.

7. Immersion bilingual education. The Immersion Bilingual Education is for students to become bilingual and bicultural without the loss of academic achievement. There are four sub-types under Immersion Bilingual Education: Early total immersion, early partial immersion, delayed immersion, and late immersion. These four sub-types of the program relate to the age (early, middle, and later) of the students starting the program, and the amount of time spent in the program (partial or total). This type of bilingual education has been practised in Canada with programs called early total, early partial, delayed immersion, and late immersion. This immersion bilingual education has received some success, due to six important features:

1. Immersion in Canada aims at bilingualism in two majority languages, such as French and English;
2. Immersion bilingual education is not compulsory but optional;
3. Home language is allowed for one to one and half years as the classroom language of instruction;
4. Teachers are bilinguals with both French and English qualifications;
5. The aim of classroom language communication is to be meaningful, authentic and relevant to children's needs;
6. Most students are monolingual when starting the immersion program. They start the program with a similar lack of experience of the second language. Therefore, the students have high self-esteem and motivation because no other students possess a higher level of language skills.

8 Maintenance and heritage language bilingual education. This education program occurs where the language minority children use their native, ethnic, home, or heritage language as a medium of instruction with the goal of full bilingualism. The six main features of this type of bilingual education are as follows:

1. Most of the time, students will come from a minority language home, but at the same time the minority language may be the majority language of the local community (for example, Spanish language is used in Spanish communities in some cities in the US). The Heritage Language Education

program teaches students through a heritage language, and includes not only heritage language students but also other native students (whose first language is English);

2. Parents have the choice to send their children to either mainstream schools or to heritage bilingual schools;
3. The language of minority students is used for half or more of the curriculum time;
4. Concepts are not retaught in the other language. Transfer in the two languages is encouraged;
5. The justification given to the heritage languages is that the minority language is easily lost, while the majority language is easily maintained. Thus, bilingualism is obtained from the initial concentration on the minority language at school. In the later stages of elementary school, attention is focused more on majority language development;
6. Heritage language schools are mostly elementary schools. However in other places, the heritage schools start at kindergarten (in Hawai'i) or at the end of secondary school (in Wales).

In the USA, this type of program is also called Maintenance Bilingual Education, or Developmental Maintenance Bilingual Education. Maintenance and Heritage Language Bilingual Education has been practised in countries like the US for Navajo and Spanish languages, Canada for the Ukrainian language, and New Zealand for the Maori language.

9. Dual Language Immersion. The aim of this type of bilingual education is language separation and compartmentalisation. There is only one language used in each period of instruction. Language boundaries are established in terms of time, curriculum content and teaching. Regarding time, a decision is made about when to teach through language on alternative days, or different lessons may use different languages. In regards to the curriculum content, some lessons are taught using one language and the other lessons are taught in another language. With regards to teaching, teachers are not allowed to switch languages when teaching in a particular language.

In the USA, Dual Language or Bilingual Education has the following features:

1. A non-English language (that is a minority language) is used for at least 50% of instruction that lasts for up to six years;
2. In each period of instruction, only one language is normally used. Instruction must be adjusted to the students' language level, but must be challenging, empowering and enabling. Language is learnt primarily through content;
3. Both English and non-English speakers are present in approximately balanced numbers and integrated for most content instruction. The English and non-English speakers are integrated in all lessons.

Having a mission to produce bilingual, biliterate and multicultural children, this type of bilingual education has five major goals:

1. High levels of proficiency in students' first language and a second language;
2. Reading and writing at grade level in both languages;
3. Academic achievement at, or above level (e.g. mathematics, science, social studies);
4. Positive intercultural (multicultural) attitudes and behaviours; and
5. Communities and society to benefit from having citizens who are bilingual and biliterate, who are positive towards people with different cultural backgrounds, and who can meet national needs for language competence and a more peaceful coexistence with peoples of other nations.

In order to meet the mission, a variety of practices are utilized, such as the following:

1. The two languages of the school are given equal status in the school;
2. The school ethos is bilingual;
3. In some Dual Language bilingual schools, the two languages are taught as languages (sometimes called language arts instruction). In some other Dual Language bilingual schools, the two languages are used as the medium of instruction in order to ensure bilingual development;
4. Staff in dual language classrooms are often bilingual; and
5. The length of the Dual Language bilingual program needs to be longer rather than shorter. Such a program for two or three grades is insufficient.

This type of bilingual education, also known as Two-Way Schools, Two-Way Immersion, Two-Way Bilingual Education, Dual Language Education, Bilingual Immersion, Spanish Immersion, Double Immersion, and Interlocking Education, has been practised in some countries like the USA (mainly elementary schools).

10. Bilingual education in majority languages

Another type of bilingual education program refers to learning a second language by students living in Asian and other countries who are originally already bilingual or multi-lingual (see Table 2.2). Thus, a second language is added as a medium of instruction to the original strong indigenous or national language. Therefore this means that some curriculum content is learnt through students' second language. This type of bilingual education has been practised in countries like Luxembourg and some European Schools.

Baker (2011) summarised the types of bilingual education that have been presented earlier. Types of bilingual education are presented in Table 2.2.

Table 2.2
Type of Bilingual Education

MONOLINGUAL FORMS OF EDUCATION FOR BILINGUALS				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
MAINSTREAMING/ SUBMERSION (Structured Immersion)	Language Minority	Majority Language	Assimilation/ Subtractive	Monolingualism
MAINSTREAMING/ SUBMERSION with Withdrawal Classes/ Sheltered English/Content-based ESL	Language Minority	Majority Language with 'Pull-out' L2 lessons	Assimilation/ Subtractive	Monolingualism
SEGREGATIONIST	Language Minority	Minority Language	Apartheid	Monolingualism
WEAK FORMS OF BILINGUAL EDUCATION FOR BILINGUALS				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
TRANSITIONAL	Language Minority	Moves from minority to majority language	Assimilation/ Subtractive	Relative Monolingualism
MAINSTREAM with Foreign Language Teaching	Language Minority	Majority language with L2/FL lessons	Limited Enrichment	Limited Bilingualism
SEPARATIST	Language Minority	Minority Language (out	Detachment/ Autonomy	Limited Bilingualism

of choice)

STRONG FORMS OF BILINGUAL EDUCATION FOR BILINGUALISM AND BILITERACY				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
IMMERSION	Language Majority	Bilingual with initial emphasis on L2	Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
MAINTENANCE/HERITAGE LANGUAGE	Language Minority	Bilingual with emphasis on L1	Maintenance, Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
TWO WAY/ DUAL LANGUAGE	Mixed Language Minority & Majority	Minority and Majority	Maintenance, Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
MAINSTREAM BILINGUAL	Language Majority	Two Majority Languages Pluralism	Maintenance, & Biliteracy and Enrichment. Additive	Bilingualism

Note: L1=First Language; L2=Second Language; FL=Foreign Language

Source: (Baker, 2011, pp. 209-210; and Ovando & Combs, 2012, pp. 35-44)

Considering the ten models of Bilingual Education above, this study would suggest that bilingual education in Aceh Province falls into the last model which is model number 10, due to the fact that the Acehnese type of bilingual education program refers to learning a foreign language by students living in Aceh Province of Indonesia when they have known a local language, such as Acehnese, Bahasa Jamee, Bahasa Tamiang and another nine local languages, and they have known Bahasa Indonesian as the national language. Thus, a foreign language is added as a medium of instruction to the national language. Therefore this means that some curriculum content is learnt through students' foreign language, in this case English.

Brief History of Bilingual Education Worldwide

Historically, bilingual education in the world has existed for as long as a thousand years (Mackey, 1978) when human society was formed and then continued through from the ancient world to the renaissance and to today's modern world (Lewis, 1977, cited in Baker 2011). Bilingual education in the world, for example, in the USA, Sweden, England, and Canada, has been associated with each country's historical context of immigration, as well as political movement (Baker, 2011). Bilingual education for some other countries like Wales and Ireland has been linked with the movements in language rights and self-government (Jones & Martin-Jones, 2004). In

Japan, it has been connected to a movement from monolingual ideology to internationalism (Maher, 1997). In Southeast Asia (such as in Indonesia, Malaysia, Brunei Darussalam, and Singapore), bilingual education has been related to the birth of ASEAN and English as a *lingua franca* in ASEAN (Kirkpatrick, 2012; Kirkpatrick & Sussex, 2012).

Well-known contexts for bilingual education include the USA and Canada. Bilingual education in the USA was believed to have started when immigrants from South America, Italy, Germany, Holland, France, Poland, Czechoslovakia, Ireland, Wales, Spain, China, and other countries, came to the USA (Baker 2011). Putting faith in the country's promise of liberty and a better life, the immigrants entered the country together with their heritage, languages and cultures. It did not take long before the immigrants were assimilated and schools for their children were set up, and it appeared that a competition between public and private schools that later motivated an openness for the language of the immigrants occurred. However, bilingual education in the USA received a setback when the country prohibited the teaching of certain languages due to an anti-German sentiment, related to the First World War. Bilingual education in the country resumed in 1963, when the first modern dual school was established in South Florida (Baker 2006). Since that time, bilingual education has become considerably accepted after a lawsuit to have it formally implemented.

The following is a summarised history of bilingual education in the USA based on work by Baker (2006, 2011).

Table 2.3
Brief History of Bilingual Education in the USA

Year	US Legislation/Litigation affecting Bilingual Education	Implication
1906	Nationality Act passed	First legislation requiring immigrants to speak English to become naturalized.
1923	<i>Meyer v. Nebraska</i> ruling the US Supreme Court	The ruling outlawed, as an unconstitutional infringement of individual liberties, arbitrary restrictions on teaching languages other than English. Proficiency in a foreign language was also constitutional.
1950	Amendments to the Nationality Act	English literacy required for naturalization.
1954	<i>Brown v. Board of Education</i>	Segregated education based on race made unconstitutional.
1958	National Defense Education Act	The first federal legislation to promote foreign language learning.
1965	Immigration and National Act	The Act eliminated racial criteria for admission,

		expanding immigration especially from Asia and Latin America. The Act also emphasized the goal of ‘family unification’ over occupational skills. This encouraged increased immigration by Mexicans in particular.
1965	Elementary and Secondary Education Act (ESEA)	Funds granted to meet the needs of ‘educationally deprived children’.
1968	Elementary and Secondary Education Act (ESEA) amendment: The Bilingual Education Act, Title VII	Provided funding to establish bilingual programs for students who did not speak English and who were economically poor.
1974	<i>Lau v. Nichols</i>	Established that language programs for language minorities not proficient in English were necessary to provide equal educational opportunities.
1974	Equal Educational Opportunity Act (EEOA)	Codified the <i>Lau v. Nichols</i> decision, requiring every school district to take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs.
1974	Reauthorization of Bilingual Education Act Title VII of ESEA	Native-language instruction was required for the first time as a condition for receiving bilingual education grants. Bilingual Education was refined as transitional (TBE).
1975	Lau Remedies	Informal guideline on schools’ obligations toward LEP students. This required the provision of bilingual education in districts where the civil rights of such students had been violated.
1976	<i>Keyes v. School District no. 1, Denver, Colorado</i>	Established bilingual education as compatible with desegregation.
1978	Reauthorization of Bilingual Education Act Title VII of ESEA	A new restriction was introduced. Grants could support native-language instruction only to the extent necessary to allow a child to achieve competence in the English language. Funding was thus restricted to TBE; maintenance programs were now ineligible for funding. The term ‘Limited English Proficient’ (LEP) introduced, replacing LES (Limited English Speaking).
1980-81	Lau Regulations	The Carter Administration attempted to formalize the Lau Remedies, requiring bilingual instruction for LEP students where feasible. The Reagan Administration subsequently withdrew the proposal, leaving uncertainty about schools’ obligations on this area.
1981	<i>Castañeda v. Pickard</i>	An Appeal court decision established a three-part test to determine whether schools were taking “appropriate action” under the 1974 Equal Educational Opportunity Act. Programs for LEP students (bilingual or otherwise) must be: (1) based on sound educational theory, (2) implemented with adequate resources, and (3) evaluated and proven effective.
1983	US English Movement launched	Debates about the dominant place of English in law, society and education became more prominent.
1984	Reauthorization of Bilingual Education Act Title VII of ESEA	While most funding was reserved for TBE, monies for maintenance programs were once again permitted, along with ‘special alternative’ English-only programs.
1988	Reauthorization of Bilingual Education Act Title VII of ESEA	Same as in 1984, but 25% of funding given for English-only Special Alternative Instructional (SAIP) programs.
1994	Reauthorization of Bilingual Education Act Title VII of ESEA	Full bilingual proficiency recognized as a lawful educational goal. Funded dual language programs that included English speakers and programs to support Native American languages. The quota for funding SAIP programs was lifted. The new law sought to bring LEP students into mainstream school reform efforts, making it

		more difficult for their particular needs to be ignored in policymaking.
1998	Proposition 227 passed in California	The 'Unz initiative' sought to impose severe restrictions on native-language instruction for English learners in California. Most bilingual programs dismantled, with similar measures in Arizona (2000) and Massachusetts (2002).
2002	No Child Left Behind legislation as a reauthorization of the Elementary and Secondary Act of 1965 and a repeal of the Bilingual Education Act	Schools and states encouraged to move to English-only education through mandatory high-stakes testing in English. Measures of Adequate Yearly Progress (AYP) reported for schools, school districts and states including English proficiency.

Source: Baker (2006, pp. 201-202; 2011, pp. 196-198)

With immigrants accounting for two thirds of the population, Canada is one of the most culturally and linguistically diverse countries in the world, representing a bilingual and pluralistic country (Clark, 2012). Canada has been famous for its immersion programs. The Canadian immersion program started in Quebec when it underwent a social transformation in the 1960s. Some English speakers went out of the city while others stayed and learnt French to compete in the markets. Originally, the St Lambert's immersion program did not contain the idea of bilingualism, and consisted of four aims (Roy & Galiev, 2011):

1. To allow students to be competent both in oral and written French;
2. To maintain the development of English as the first language;
3. To allow students to learn content according to their age and level of school; and
4. To aid French speaking students and English speaking students to develop understanding and respect for each other's culture and language while preserving their own culture and language (summarised from Roy & Galiev, 2011).

The success of the St. Lambert's immersion education program promoted bilingualism which was then used to transfer to other schools where English was dominant.

In Southeast Asia, English has been spread remarkably well. Its spread has been associated with the establishment of the Association of Southeast Asia Nations (ASEAN). Representing ten countries: Indonesia, Malaysia, Singapore, the Philippines, Thailand, Brunei Darussalam, Vietnam, Laos, Myanmar, and Cambodia, it was founded in 1967 to stand against communism, as communism was then seen as a threat in the region. Even though there was no agreement to use English as the language for communication among the ASEAN countries, English was chosen as the *de facto* language for communication, also known as a *lingua franca*, and the decision was considered 'natural' (Kirkpatrick, 2007, p. 137). In 1999, it became the official language for ASEAN (Kirkpatrick, 2012).

English has played an important role in the curriculum of ASEAN countries. It has been taught in schools with varying degrees and emphasis. Some countries have regarded it as a medium of instruction, and some other countries have used it as a core subject. English is the medium of instruction for all subjects in Singapore, as it is in the Philippines and Brunei Darussalam. In Malaysia, it has been the medium of instruction for 11 years (from 2002 to 2013). In Indonesia, English has been taught as a foreign language for a number of years, and it is also regarded as a medium of instruction for some levels and types of schools. In Myanmar, English is taught as a separate subject (Kirkpatrick, 2012).

In recent years, however, there has been some controversy in some ASEAN countries as to whether English is best regarded as a core subject only, or as a medium of instruction. Different countries have dealt with this controversy differently. Malaysia, which has used English as a medium of instruction for subjects like Mathematics and Science in secondary schools since 2002, reversed this decision in 2013 and returned to Bahasa Malaysia as the medium of instruction. The reversal was due to complaints that the use of English as the medium of instruction only benefitted the urban area-students, and the rural area-students were left behind (Kirkpatrick & Sussex, 2012). Similarly, Indonesia has in 2013 changed the use of English, from its position as a medium of instruction for some types and levels of schools, to teaching it only as core subject with a lesser amount of class time. The reason behind this decision was based on certain developing views that Bahasa Indonesia should be emphasized and regarded as the medium of instruction. In contrast to Malaysia, Brunei Darussalam has increased the

amount of English used as the medium of instruction, leading to its use as the medium of instruction for Primary 1, for certain parts of the curriculum (see Kirkpatrick & Sussex, 2012).

The Nature of Bilingual Education in Indonesia

What is the Position of Bahasa Indonesia?

Bahasa Indonesia (also called Indonesian) has been regarded as the Indonesian national language since *Sumpah Pemuda* declaration (The Youth Pledge) on October 28, 1928. The pledge contained three oaths, that is: (1) to acknowledge one's motherland: that is Indonesia; (2) to acknowledge one nation: that is Indonesia; and (3) to uphold one language of unity: Indonesian. Bahasa Indonesia was utilised as the national language well before the declaration of Dependence Day took place in August 17, 1945. Bahasa Indonesia has been regarded as the most important language in Indonesia even though it was not the most spoken language. The reason for choosing it as the Language of Unity was because it had been the least 'treated' language compared to Javanese which had been spoken widely in Indonesia. At that time, Javanese was a language of elite politics in Indonesia. Should Javanese be taken as a national language, it would lead to a conflict of interest amongst the Indonesians, since it would bring privilege to certain powerful community groups and it would emphasise the role of the hierarchy that culture in the language brought with it (Kirkpatrick, 2011).

As a language that unites 726 dialects spoken by more than 400 ethnic groups in Indonesia, Bahasa Indonesia has been a great success (Kirkpatrick & Sussex, 2012). Today, it is spoken by more than 200 million Indonesians (Hamied, 2011, in Kirkpatrick, 2011), and it functions well in administration, politics, and the judiciary. Bahasa Indonesia is now the main medium of instruction in educational institutions at all levels throughout the country.

Why is English in the Indonesian Curriculum?

English has been taught as a foreign language in Indonesia as a compulsory subject at secondary levels. For elementary levels, it was not compulsory, and therefore was not part of the English curriculum. However, some elementary schools have taught it as part of their local contents, due to school preferences and prestige. As local content,

it was not part of the curriculum assessment, so it was not tested for the end-of-academic-year exit examinations.

The paradigm in the Indonesian educational context was changed after 2003 in regards to a new perception towards the role of English as an international language and as the language of technology. The government of Indonesia, on behalf of the Ministry of Education, believed that the country would benefit by making English compulsory for Indonesian students at secondary level, so that students could master English as a foreign language. This decision was implemented in 2003 by the Act of the Indonesian Republic Number 20, Article 50, and Verse 3. The Act advocated that it was imperative for the Indonesia government and for local governments to establish at least one unit of English education with the international standard at all levels of education. The aim of this Act was to ensure that Indonesian citizens master English skills in order to be able to compete in different fields, such as education, information technology, communication, trade, and social culture.

Why Do They Have Bilingual Education in Indonesia?

The Act has been translated as the provision of bilingual programs in certain schools nation-wide. A number of schools were established as pilot schools, named as “International Standard School” (SBI or *Sekolah Bertaraf International*) and “Prospective International Standard School” (RSBI or *Rintisan Sekolah Bertaraf International*), while regular schools functioned normally. These schools replaced the use of Bahasa Indonesia with English. The criteria for an International Standard School are as follows (Depdiknas, 2007):

1. English is to be used as the medium of instruction for science, mathematics and core vocational subjects from year 4 of primary school and throughout junior secondary school, senior secondary school and vocational secondary school;
2. Teachers must possess the competence required to teach their subjects through English; and
3. Head teachers must possess active mastery of English (Depdiknas, 2007).

In order to be appointed as a Prospective International Standard School, certain “criteria need to be met among which is the teaching of at least two subjects in English”

(Chodidjah, 2009, p. 88). In implementing bilingual education, the Indonesian government and local governments have facilitated some innovative programs, such as ensuring bilingual teachers are actively involved with in-service training and sending the teachers to study abroad. Such programs were intended to help bilingual teachers equip themselves with the knowledge and skills needed for success in teaching in the bilingual program. In addition to that, the government also supported the schools with modern facilities in order to enable the teaching and learning bilingual programs to run as expected.

The bilingual program gained success after a couple of years of implementation. However, along the way, it also received many criticisms. Opponents of bilingual education have criticised the government for not being ready to carry out the bilingual program to the standard required in terms of school facilities, teachers, and students. While the idea of bilingual education was positively accepted, the implementation of it has been seen as immature, in term of insufficient preparation or understanding of what was needed (Davies, 2005). Bilingual schools have also been associated with luxury and prestige, due to the fact that only students from the richer families were able to enrol in them. Students from poor families could not afford it, since some bilingual schools required financial support from students (Revianur, 2013b). Another criticism focused disagreement about Bahasa Indonesia being replaced by English. A fear was raised that English would diminish students' valuing of the Indonesian language and culture (Tubagus, 2005).

Bilingual Education Program Ban in Indonesia (2013)

Due to criticism that bilingual education was seen as discriminating against a fair education, some groups brought this case to the Judicial Court (known in Indonesia as *Mahkamah Konstitusi*). With regards to this issue, the Judicial Court came to the agreement that bilingual programs had to be eliminated from the Indonesian education system because they violated the basic law of the Indonesian Republic (Year 1945) and they did not have a strong supporting background case (Revianur, 2013b). The previously bilingual schools returned to standard programs (Revianur, 2013a). This ban on bilingual education programs came into effect in January 2013.

Coinciding with the banning of bilingual education in Indonesia, a new 2013 curriculum has been launched. The primary change regarding English stated in the new curriculum was that there would be no bilingual education in the Indonesian educational system and the amount of teaching English has been drastically decreased. This change in regards to the English curriculum has a number of effects. It has been predicted that many English teachers would lose their jobs due to the omission (or marked decrease) of their teaching hours. Students who have previously benefited through bilingual educational programs may find it difficult, or unhelpful, to return to the original mode of separate English teaching classes. It would be expected to take some time for many previously-taught bilingual students to readjust. Handbooks and selected texts and curriculum books would need to be replaced with those re-written for the new curriculum and this may take some time also.

What is the future of Bilingual Education in Indonesia?

It would seem that it is not easy to see much of a future for bilingual education in Indonesia in the near future from 2013 onwards, especially with the implementation of the new curriculum where bilingual education is banned by law. However, as the trends in the Indonesian educational system are dynamic, one can hope that a revised decision could be made that benefits Indonesian students in terms of having opportunities to learn additional languages in order to enrich and prepare students for a global village of bilingualism and multilingualism.

A Review of Bilingualism and English as a Second or Foreign Language

There have been a reasonable number of studies conducted on second language acquisition, bilingualism, English as a second language, and English as a foreign language, in various contexts with different data analyses and measures. This section revisits three areas of the research that focused on: (1) Review of second language acquisition, (2) Findings on experimental and control groups relating to reading and writing, and (3) Findings on Rasch measures in second language.

Review of Second Language Acquisition

A review of second language acquisition has recently been conducted by Dixon et al. (2012) and published in the Review of Educational Research Journal. The study investigated relevant papers from 72 peer-reviewed journals that had been published from 1997 to 2011. Due to its concise and complete work, the study (pp. 5-60) outcomes are summarised for this section.

The study entitled “What we know about second language acquisition: A synthesis From Four Perspectives”. It looked at four perspectives on second language acquisition, namely; (1) Foreign language educators; (2) Child language; (3) Sociocultural; and (4) Psycholinguistics. Each perspective focuses on a different unit of analysis, context, age of learners, language domain focused on, and major questions (see Table 2.4).

Table 2.4
Difference among the Four Perspectives

Perspective	Unit of analysis	Context	Age of learners	Language domain focused on	Major questions
Foreign language (FL) educators	Student achievement or proficiency	FL classroom with little FL exposure in community	Adolescents and adults	Accent, grammatical correctness, reading ability	Developing and improving teaching methods and techniques
Child language	Child	Naturalistic, extended to teacher or student in classroom	Young children	Grammar, lexicon, extended discourse (narrative, academic language)	Identifying factors that influence child L1 development, bilingual L1 acquisition, child L2 acquisition
Sociocultural	Students and interlocutors	Naturalistic or classroom	Any	Communicative effectiveness, pragmatics	Understanding social and cultural differences, impact of identity, interpersonal interactions
Psycholinguistics	Cognitive and brain processes	Laboratory, extended to classrooms, naturalistic	Any	Grammar, lexicon, text structures	Understanding process of L2 acquisition, transfer of L1 skills to L2.

Note: L1= First language; L2=Second language

Source: Dixon, Zhao, Shin Wu, Burgess-Brigham, Gezer, & Snow, 2013, p. 8.

The review was carried out to address five questions of particular relevance to the four perspectives and their relevance to second language education. They are:

1. What are the optimal conditions for second language acquisition?
2. What are the characteristics of excellent or unsuccessful second language learners?
3. What are the characteristics of excellent or unsuccessful second language teachers?
4. What are reasonable expectations for speed and accomplishment for second language learners of different ages?
6. Has information generated by the four research perspectives influenced the formulation of educational policies for second language learners?

The findings provided in the review were based on answers to the five questions above.

Findings for Question 1: What are the optimal conditions for second language acquisition?

There was no confirmed definition on ‘optimal conditions’ for the best way to teach second language learners. Optimal conditions for acquiring a second language for different populations vary in regard to learning contexts, pedagogical goals, program setup, learner characteristics, and the interactions among these variables. The best conditions for second language learners in second language-majority contexts include higher family socioeconomic status together with parent and grandparent education, strong home literacy practices, opportunities for informal second language use, and well-designed and well implemented educational programs specifically for second language literacy instruction. Regarding this, teachers can make a positive difference by encouraging home literacy, sending home books and other literacy materials and getting the parents to read with their children in either second language or first language. Further positive help can be provided by taking them to the library, and supporting informal second language use by mixing second language learners with first language speakers, as well as promoting integrated curricular activities and ensuring that there is

sufficient time to implement educational programs and lesson plans that follow research-tested designs in their schools

Optimal conditions for second language learners in a foreign language setting were stated as explicit instruction helps students, especially in learning grammar, and using academic content to teach the second language may be beneficial to building vocabulary in a second language. However, it was also stated that the intensity of second language instruction seemed to make little or no difference in second language learning performance.

Findings of Question 2: What are the characteristics of excellent or unsuccessful second language learners?

The following findings summarise the answers to Question 2. A positive aptitude and motivation of second language learners contribute to the largest positive difference in second language outcomes. Better first language skills, lower second language anxiety, and gender (females do better at language) contribute to positive second language outcomes. Verbal ability is more important to older second language learners than to younger second language learners. First language literacy skills are reasonable predictors of second language literacy skills and the outcomes for foreign language students and for learners in second language-majority settings are influenced by different factors.

Findings of Question 3: What are the characteristic of excellent or unsuccessful second language teachers?

Competent second language teachers possess adequate proficiency in the target language. They desire to teach well and there is a positive correlation with self-efficacy, intellectual excitement, and teacher reflectivity. Good teachers have good classroom organization where instruction is clear and well planned, and students know what to expect. Good second language teachers have, at least, some of the students' first language proficiency and they know when and how to use it.

Findings of Question 4: What are the reasonable expectations for speed and accomplishment for second language learners of different ages?

For learners in second language-majority contexts, young second language learners are likely to be successful at second language only after several years. These second language learners can reach native-like proficiency in oral fluency, vocabulary, and literacy tasks, even when they start after the age 12, but for the latter, this proficiency does not necessarily include grammar and pronunciation. In the USA context, second language learners who arrived as teenagers succeeded academically in the second language, given appropriate input, instruction, and motivation.

Second language learners' first language skill affects their rate of second language acquisition. Learners, whose first language was more distant to the second language (such as Korean and English), took longer to acquire competence. For second language learners in foreign language contexts, the canard that says 'younger is the better' should be rejected and, holding hours of instruction constant, older learners performed better on the proficiency measures but, of course, an early start could be beneficial for providing more hours of input.

Findings of Question 5: Has information generated by the four research perspectives influenced the formulation of educational policies for second language learners?

In the USA, the Act of No Child Left Behind (NCLB) had unforeseen consequences for second language learners and second language education. The Act was not supported by the four perspectives. USA state policies indicated that one year was considered sufficient for second language learners to gain English proficiency in a mainstream classroom. This, however, was not supported by the research from the four perspectives. Two-way bilingual education was effective at the pre-school level and some two-way programs with voluntary enrolment were effective at the pre-kindergarten to 8 levels.

There was no national policy (in the No Child Left Behind Act) in regard to foreign language teaching or second language learning. USA policy did not incorporate the research findings of any of the four bodies of research work on English in the second language learner context or in any other language.

(Summarised by Khairiah Syahabuddin from Dixon, Zhao, Shin Wu, Burgess-Brigham, Gezer, & Snow, 2013, pp. 5-60, from the Review of Educational Research Journal)

Findings on Experimental and Control Studies for Reading and Writing, and Attitude and Behaviour

There have been a number of studies using experimental and control groups on the four sub-skills of English proficiency, namely, listening, speaking, reading, and writing. This section, however, only revisits studies on reading and writing due to the fact that they were the core interest of the present study. The studies on reading and writing are revisited prior to the studies on attitude and behaviour.

There have been a number of recent experimental and control group studies related to reading comprehension. For example, Alshumaimeri and Almasri (2012) conducted an experimental and control study on the effect of using WebQuests on Reading Comprehension performance. The respondents for the experimental group were 42 male Arab students and they received traditional teaching together with WebQuest additional activities. The respondents for the control group were 41 male Arab students and they received only traditional teaching. The performance data were analyzed using SPSS and t-tests. The results indicated that the WebQuests had potential in promoting reading comprehension.

Buyuktaskapu (2012) carried out a study to examine the effect of family support during pre-reading training on reading success in a Turkish primary school. There were 25 first graders participating in the Family Supported Pre-Reading and they were taken as the experimental group. The control group consisted of 25 first graders who attended nursery class. Using various true-score non-linear scales of reading and writing to analyze the data, the study revealed that children attending the Family Supported Pre-Reading Program had more reading success than children attending nursery classes.

Nakanishi and Ueda (2011) explored the effect of extensive reading and shadowing on reading comprehension. Shadowing has been defined as a task of listening in which students track the target speech and repeat it as immediately and quickly as possible without referring to a text. The study used 89 first-year Japanese university students to comprise two experimental and one control groups. Data were collected using the Secondary Level English Proficiency Test of students' reading comprehension and analyzed with ANOVA to compare groups. The results showed that there was no statistically significant dissimilarity among groups.

Sahin (2011) analyzed the effect of the way Turkish students read a reading comprehension text, that is, through page scrolling or page-by-page reading, in relation to the students' reading comprehension performance. Forty-six fourth grade students who were in one experimental group and one control group were the participants of the study. T-tests were used to analyze the data. The result suggested that there were no significant differences between the comprehension scores of students using the scrolling reading method and the scores of students using the page-by-page reading method.

Bolukbas, Keskin, and Polat (2011) conducted a study on the effectiveness of cooperative learning on reading comprehension skills. The respondents were 40 Turkish university level students, consisting of 20 students who used a cooperative learning technique as the experimental group and 20 students taught using a traditional teaching model. SPSS 11.00 was used to analyze the data. The results confirmed that cooperative learning is more effective in improving learners' reading comprehension skills.

Novel studies on writing have also been conducted. Tajeddin and Daraee (2013) studied the effect of form-focused and non-form focused tasks on vocabulary learning through written forms of English as a foreign language. This study used 50 Iranian students who learnt English as a foreign language, consisting of two experimental groups and one control group. Data were analyzed using ANOVA and resulted in various conclusions, indicating that task effectiveness depended upon the level of involvement and the task's influence and complexity.

Tabatabaei and Assefi (2012) explored the effects of a portfolio assessment technique as a teaching, learning, and evaluation instrument on EFL learners' writing performance. Forty Iranian students majoring in English participated in the study and were divided into an experimental and a control group. The portfolio assessment was assigned to the experimental group and the traditional approach of writing assessments was assigned to the control group. Using the TOEFL and IELTS tests, the result indicated that the students with portfolio assessment outperformed their counterparts in both writing performance and writing sub-skills.

Hosseini (2012) conducted a study to investigate the effectiveness of asynchronous computer-mediated corrective feedback, both explicit and implicit, on increasing the prepositions' correct use. The respondents were 45 Iranian elementary

EFL learners, who were assigned to two experimental groups and one control group. The study used a one-way ANOVA and concluded that the students who received explicit corrective feedback outperformed the students who only received implicit corrective feedback and those who received no corrective feedback.

Marashi and Dadari (2012) explored the impact of task-based writing on Iranians' learning ESL and their writing performance and creativity. The participants were 56 students grouped into experimental and control groups. Using the Abedi-Schumacher Creativity Test and Cambridge General Mark Scheme for Writing, the results showed that learners benefited significantly from task-based writing and creativity.

There has been a reasonable number of studies conducted on attitude and behaviour with regard to learning English as a second or foreign language (see Alarcon, 2011; Ali, Mukundan, Baki, & Ayub, 2012; Gebhard, 2012). However, only a few studies have been conducted using experimental and control groups. Soleimani, Mainnzadeh, Kassaian, and Ketabi (2012) aimed their study at the effect of instruction based on Multiple Intelligence Theory with regard to attitude and learning of general English. The study recruited 61 Iranian female and male students who were randomly divided into one experimental class and one control class. ANCOVA and independent sample t-tests were used to analyze the data. The results indicated that students taught on the basis of Multiple Intelligence Theory exceeded students who were taught traditionally. They outperformed both in general and sub-skills of learning English. The results also revealed that students who were taught on the basis of Multiple Intelligence Theory significantly improved their attitude toward learning English.

Mekheimer (2012) conducted research on the effect of using Blackboard technology (e-learning) and online dictionaries with English as a foreign language, based on students' translating skill and attitude towards learning English. The study involved 83 male Saudi Arabian university students who were assigned to experimental and control groups. The data were analyzed with t-tests and showed students' translation skills improved together with improvements in positive attitudes toward translation.

Merisuo-Storm (2007) investigated students' attitudes, with regard to foreign language learning and the development of literacy skills in bilingual education. The research consisted of two studies carried out within two years. The first study used 78 students who were divided into three experimental groups, and 58 respondents who joined three control groups. The second study involved 70 students in three experimental groups and 75 students in three control groups. Data were analyzed using the SPSS program and the results showed that students' literacy skills in the bilingual classes were significantly better than those in the monolingual classes. In addition to that, the results indicated that the students in bilingual classes showed significantly more positive attitudes towards foreign language learning than the students in the monolingual classes.

Yoshimura (2006) carried out a study on whether manipulating foreknowledge (schemata) of output tasks leads to differences in students' reading behaviour, text comprehension, and noticing of long forms. Participants for the study were 57 Japanese university students who sat for three different experimental tests. The data were analysed using ANOVA and the results suggested that foreknowledge of output tasks influenced the students' behaviour. It also revealed that the students performed more translation into their first language, expressed more interrelation between their target language and inter-language, and engaged more with language form when foreknowledge manipulation of output tasks was used.

Despite the scarcity of gender studies in EFL research, in terms of pretest-posttest studies, there are studies focusing on gender in relation to students' reading comprehension and writing achievement, and attitude and behaviour. Rahmani and Sadeghi (2011) conducted a study to examine the process and product effect of note-taking strategy on reading comprehension and written material, with gender as a moderating variable. The respondents were 108 undergraduate Iranian EFL students who were assigned into two groups. The experimental group received training on how to take notes and to use graphic organisers as guide. The control group did not receive any instruction. The results of the Two Way ANOVA suggested that the experimental group performed significantly better, remembered more important ideas, and better identified the relationship between ideas. There was no statistically significant effect on gender on students' performance in the comprehension and retention tests. A study to

investigate the effect of multiple intelligence strategies, covering logical-mathematical intelligence and interpersonal intelligence and interpersonal intelligence was conducted by Abdelrahman and Jallad (2008). The study took ninth graders for its respondents who comprised of two male sections and two female sections. Using the Two Way ANOVA, the study showed that there was a significant difference in the students' reading comprehension due to the teaching strategies in favour of the experimental group. There was no significant difference in the students' reading comprehension in regards to the students' gender.

Jafari and Ansari (2012) conducted a research on students' writing achievement. They investigated the effect of group work on Iranian EFL learners' writing accuracy and the effect of gender on text production. The respondents were divided into one experimental group and one control group. The experimental group wrote collaboratively while the control group wrote individually on four essay writing sessions with the same topics and genre. The results revealed that the experimental group outperformed the control group, and females outperformed males. Sadeghi and Sharifi (2013) investigated the effect of post-teaching activity, comprising game, narrative writing, role-play, and speaking tasks on vocabulary gain of EFL learners. The sample of study was 111 elementary female and male adult EFL learners. The respondents were grouped into four experimental groups for females and four experimental groups for males as well as two control groups for each gender. The results of the Two-Way ANOVA indicated that there were statistically significant main effects for vocabulary learning across different activities that lead to the higher vocabulary gain. Female learners outperformed male learners.

Murad Sani and Zain (2011) conducted a study on attitude and behaviour. They investigated the relationships among second language reading attitudes, reading self-efficacy, and reading ability, as well as gender differences in a non-supportive ESL setting. The respondents were 218 students who completed a translated version of Middle/Secondary Reading Attitude Survey and English reading measure. The result stated that the students' second language reading ability was average, their second language reading self-efficacy was low, and the respondents' attitudes were not positive. There was no significant gender difference in self-efficacy. However, female students showed more promising attitudes and comprehended significantly better.

Arshad and Ali (2013) conducted a study to explore whether gender differences effect on learning English in a culture of Pakistan. A self-reporting questionnaire based on a 5-point Likert Scale was used to collect the main data. The participants were students of 'immediate class' and the data were analysed with SPSS. This suggested that social factors have a great effect and impact on learning the English language.

Findings Using Rasch Measures and Control and Experimental Groups

While some research has been done with Rasch-created linear measures on reading, writing, and attitude and behaviour (see Knoch, 2010; Koh, 2008; Lim, 2011; Metsämuuronen, Svedlin, & Ilic, 2012; Wang, Kim, Bong, & Ahn, 2012), research with Rasch measures conducted using experimental and control groups for English as a second language remains scarce. Waugh, Bowering, and Torok (2005) created linear Rasch scales to measure reading comprehension, and attitude and behaviour for Thai English as a second language students. The study explored the use of various genres to improve the teaching and learning of English reading. The participants were 300 grade 7 students who were taught English as a second language through genre-based methods, which were called Expository Genre, Narrative Genre, and Journalistic Genre, and were compared to students taught through a traditional communicative method. The results showed that students taught through the genre method outperformed students taught in the traditional way on reading comprehension and attitude and behaviour. In another study in Thailand, Waugh, Bowering, and Chayarathee (2005) studied grade 6 Thai students who were taught through the cooperative learning method and compared their performance with students taught through a traditional teaching method. The study showed that students taught through a cooperative method did better in reading comprehension, and had better attitudes and behaviour with regard to learning English than those taught with the traditional communicative method.

Research on English Reading Comprehension and English Writing as well as on Attitude and Behaviour have been conducted during this decade (2003-2013). Most of the research studies have drawn conclusions that bilingual students outperformed their counterparts. However, there was no study conducted to cover all the three variables, and in Aceh Province with Acehnese context, as the present study did. Therefore, it was considered a need to conduct the present study in order to see how Acehnese students

dealt with bilingual education as part of their challenge in education rehabilitation and revitalisation after the Tsunami 2004 disaster.

Gaps in Indonesian Research: Bilingualism *versus* Monolingualism

There appear to be four main gaps in the research literature in Indonesia (and especially Aceh) in regard to bilingualism and bilingualism education. One, while there is much anecdotal evidence in Aceh that bilingual schools are better than monolingual schools and that students taught bilingually perform better than students taught monolingually, there is no research evidence for this from Indonesia and Aceh. There is evidence from other countries supporting this, but not from Aceh itself. The judicial decision to ban bilingual teaching in Indonesia appears to have been made without any research evidence on the supposed benefits of bilingual teaching and learning. Secondly, there does not appear to have been any studies in Aceh or Indonesia involving experimental groups (with bilinguals) compared to control groups (with monolinguals). Thirdly, no relevant studies have been conducted in Indonesia (or Aceh) using modern measurement methods such as Rasch models to create linear scales of achievement and behaviour in regard to research relating to bilingually taught students. All the studies, whether relevant or not, have used True Score Theory (percentage scores) which are non-linear scores and items have not been created on a scale from easy to hard as is expected of linear measures. Fourthly, there does not appear to have been any research in Aceh which actually asks bilingually taught students what they think about bilingualism and bilingual teaching. No research appears to have been done to ask these students what benefits they think they are getting in learning English bilingually. The present research study aims to address these gaps (or deficiencies), at least partially.

The result of the present study would inform the education policy makers in Aceh about the current reality in Aceh that bilingual education was still needed, apart from the banning of bilingual education throughout Indonesia and the changes made to the current Curriculum 2013 in which the percentage of English lessons taught in primary and secondary schools in Indonesia schools was reduced. It would also suggest that the Acehnese government should reconsider the banning of bilingual education in Aceh Province and reestablish state international standard schools and state candidates

for international standard schools that have been banned due to the banning of bilingual education in Indonesia. As a province with a privilege to manage its own education policy, Aceh Province has the power to do this.

The next chapter explains the conceptual framework of bilingual and monolingual teaching, measurement, and the variables used in the present study.

CHAPTER THREE

CONCEPTUAL FRAMEWORK FOR BILINGUAL/MONOLINGUAL TEACHING, MEASUREMENT, AND THE VARIABLES

This chapter presents the conceptual framework behind bilingual and monolingual teaching in Aceh province, Indonesia, and explains why bilingual education is expected to produce superior achievements over monolingual education after a two month experiment. It also presents the three dependent variables of the study, that is English Reading Comprehension, English Writing and Attitude and Behaviour with regard to learning English, and explains the conceptual structure of each variable. Then there is an explanation of the problems of True Score Theory measurement and why it is not used in the present study. A better alternative – Rasch measurement – is explained, along with the output from one of the best Rasch computer programs, RUMM 2030 (Andrich, et al., 2010). RUMM 2030 produces some excellent graphics and tabular data to support the creation of linear unidimensional scales.

Conceptual Framework for Bilingual and Monolingual Teaching in Banda Aceh

Bilingual programmes and monolingual programmes in Aceh share similar characteristics. The characteristics are: similar entry knowledge, the same curriculum, the same amount of time for classroom teaching and learning, similar classroom English activities, similar outdoor English activities, similar time spent on homework, similar textbooks used, similar teachers' ability and similar number of other subjects to studied within one academic year.

The only difference between the two program types is the medium of instruction. Bilingual programs use a majority of English in English lessons (about 50% English and 50% Bahasa Indonesia), while monolingual programs use majority Bahasa Indonesia with a little amount of English in the English teaching-learning processes, or even, use Bahasa Indonesia combined with a local language (in this case, Acehnese)

with a little amount of English, especially for students in remote areas where the local language (Acehnese) is more convenient to use in teaching any subject, including English. This difference in language use, Bahasa Indonesia or English, may contribute to different outcomes for English learning for both types of students. In terms of students' ability in three variables of this research, that is, English Reading Comprehension, English Writing and Attitude and Behaviour with regard to learning English over a two month time difference, it is expected that bilingual students will produce better results than the monolingual students in regard to English Reading Comprehension, English Writing, and Attitude and Behaviour.

In regard to English Reading Comprehension, it would be reasonable to expect that bilingual students have a number of abilities in order to help them achieve higher scores in Reading Comprehension tests than their monolingual peers. First, bilingual students have more advanced phonological awareness than monolingual students (Bialystok & Herman, 1999). Not only is phonological awareness the most important of the metalinguistic skills that are basic to reading comprehension, but it also has repeatedly and independently been shown to underlie access to literacy and to ensure progress into fluent reading (Bialystok & Herman, 1999).

Second, being in a classroom with a second language instruction, bilingual students get benefits in interaction and communication with adult teachers. In this interaction and communication with teachers, bilingual students get models and response from their teacher to the students' meaning. It is strongly believed that when dealing with two languages, bilingual students tend to predict and hypothesise second language use and forms. They can expect, for example, word order, sentence structures, and a number of formal features of the utterances they use and hear (Cromdal, 1999). In line with this, the teacher would serve as model and could respond to language use and forms that the bilingual students have predicted and hypothesised. Being able to work out the predictions and hypotheses, the bilingual students also do self-correction. Self-correction, together with predictions, hypotheses and teacher's emphasis would enhance language use and forms, which bilingual students need and are capable of implementing.

In addition to that, the language interaction between teachers and bilingual students in a classroom would help students' auditory and visual perception and memory (Merisuo-Storm, 2007), and this is probably connected with learning a second or foreign language. In learning through a foreign language, bilingual students need to pay more attention to the second language. As a consequence, the students learn to listen with fuller attention than otherwise. They need to distinguish sounds, intonation, and stresses. This leads students to develop their auditory skills and visual perceptions. By also being able to know things in two names, bilingual students tend to memorise foreign words. Memorising foreign words helps develop the bilingual students' memories. These abilities are believed to help bilingual students to achieve better scores in Reading Comprehension than their monolingual counterparts.

In relation to English Writing, since bilingual students have literacy experience in both languages, it is expected that bilingual students would perform better in writing tests than monolingual students. Written language is a symbolic system in which letters represent sounds to indicate the phonological structure of a word, although this occurs less regularly in English than in Bahasa Indonesia. The letters are symbolic because they have no meaning and do not resemble the sound they represent (Bialystock, 1997). The bilingual classroom situations, in which students experience two different symbols for almost every subject in the environment, have enhanced the students' symbolic development. Students become familiar with many phrases and are capable of recognising written words. Being in a bilingual classroom with interaction and communication with a teacher and adopting the teacher's model, bilingual students can develop spelling skills that reduce spelling errors. They also tend to be able to control their grammar over anomalous sentences (Cromdal, 1999). These faculties are believed to help bilingual students to achieve better scores in writing tests than monolingual students.

In line with the learning of Reading Comprehension and Writing, it is important to note that bilingual students, especially those with low English proficiency might find it difficult to cope with English alone. In such condition, they might need to depend on the use of their first language to help them. When their proficiency increases, their dependence on their first language decreases.

Regarding the Attitude and Behaviour tests, it was expected that students in bilingual classes would have a significantly more positive attitude and behaviour with regard to studying English than their peers in the monolingual classes. It was expected that as the bilingual students showed their advanced ability in reading comprehension and writing tests, there would be a strong link to their confidence in reading and writing skills. The students who wrote and read well were the most successful in acquiring English. Consequently, the bilingual students had a significantly more positive attitude and behaviour with regard to reading and writing than the monolingual students. This is due to the expectation that bilingual students enjoy using English language more than monolingual students and that they are proud of their knowledge of and abilities in a foreign language. Being able to interact and communicate using a language that they like can awaken the students' interest in that language and, as a result, make them eager to further their command of English, and further improve their success in learning the subject matter as well. Positive attitudes towards language learning can raise motivation and help language learning.

To sum up, it would seem reasonable to expect that the bilingual situation in which a second language is used as the medium of instruction could affect not only students' ability of linguistic processing, but also their mental representation of the language. The explanation probably may be found in the students' ability to select and apply linguistic knowledge at will and thus, these abilities would keep them motivated. This would be expected to lead bilingual students to achieve better scores in reading comprehension, and writing tests, and to develop better attitudes and behaviour with regard to learning English than monolingual students.

It is expected that monolingual students would not benefit from their English monolingual program in regard to English Reading Comprehension, English Writing, and Attitude and Behaviour, as much as the bilingual students after two months of learning English reading comprehension and English writing. It is predicted that monolingual students would show less improvement in English, that is, English Reading Comprehension, English Writing and Attitude and Behaviour results, compared to those in the bilingual programme. By using Bahasa Indonesia or a local ethnic language (Acehnese), monolinguals do not have such a good exposure to English as those taught bilingually, nor do they spend much time on English tasks despite their lessons being of

the same duration. They do not employ metacognitive skills in English reading and writing as effectively, due to the expectation that they are not under pressure to do so. By having an inadequate English vocabulary, students find that English is not as important to them and not so challenging. This may lead to a poorer result in both English Reading Comprehension and English Writing scores. Feeling not so motivated towards learning English is expected to lead to a less positive feeling towards English. Therefore, it is predicted that monolinguals would show less improvement in their attitude and behaviour with regard to learning English after a two month time, compared to those taught bilingually.

Measurement

True Score Theory Measurement

True Score Theory Measurement, which is also called Classical Test Theory, is a theory that is based on the total correct score from a set of questions, which are not ordered by difficulty. This measurement is usually designed for particular subjects, such as English, Physics, History and so on (see Chapman, 2006; Keats, 1997). The model behind this measurement is that a 'true score' and a random error score from the particular set of questions and persons make the total correct score. This measurement is commonly considered to have at least six problems: First, non-linearity; second, multi-dimensional with 'noise'; third, item difficulties not ordered; fourth, person 'measures' and item difficulties are not ordered on the same scale; fifth, the 'measures' are item-content test dependent; and sixth, the 'measures' from different tests, even on the same topic, cannot be validly added or linked onto a single scale (see Michel, 1990; Smith, 1996; Waugh & Chapman, 2005; Wright, 1996).

Total scores are not considered linear due to the fact that equal differences between them do not represent equal amounts of understanding in whatever is being measured. In other words, the difference between 20% and 30%, and between 40% and 50%, and between 60% and 70%, for example, do not represent the same amount of understanding, or knowledge, as the case may be. While a total correct score usually represents a variety of knowledge, skills and understanding, this total score cannot accurately predict a person's response to questions in tests that consist of a set of questions, without considerable further analysis. It is also believed that total correct

scores contain a lot of ‘noise’ because there are no tests usually done in regard to unidimensionality, and also there are no tests usually done on consistency of student responses on an order of difficulty linear scale, that is, question difficulties are ordered on a continuum from easy to hard. For this study, the order of difficulty used is ‘easy’, ‘hard’ and ‘harder’.

Rasch Measurement

Contrary to the True Score Theory Measurement, in Rasch measurement, items are ordered from easy to hard on a continuum and their difficulties are calculated on a linear scale, which is also known as a log odds scale (see Andrich, 1988a; Andrich, 1989; Rasch, 1960). The person measures are calculated on the same linear scale. It is important to understand that when the data fit a Rasch measurement model, the differences between the person measures and the item difficulties can be calibrated together in such a way that they are freed from the distributional properties of the incidental parameter, because of the mathematics involved in the measurement model. This means that ‘scale-free’ measures and ‘sample-free’ item difficulties can be estimated with the creation of a mathematically objective linear scale with standard units. The standard units, called logits, are the log odds of successfully answering the items. Rasch measurement has been applied to many variables in education and educational psychology (see Waugh, 2003, 2005, 2010a; Waugh, 2010b).

The Simple Logistic Model of Rasch

The simplest Rasch measurement model for creating a linear scale was developed by the Dane, Georg Rasch (1901-1980) and originally published in 1960, with expanded editions later (Rasch, 1980, 2010). The Simple Logistic Model (SLM) of Rasch has two parameters: (1) representing a measure for each person on a variable and; (2) representing the difficulty for each item, although this is sometimes called the one-parameter model in the literature. The following are the characteristics of the Simple Logistic Model (SLM) of Rasch measurement (Rasch, 1960; Waugh, 2007).

- a. Items are designed to be conceptually ordered by difficulty along an increasing continuum from easy to harder for the variable being measured. For the purpose of explanation here, only three items are ordered from easy to medium to hard.

- b. In designing the items, one keeps in mind that person measures of the variable are conceptualised as being ordered along the continuum from low to high according to certain conditions. The conditions in this example are that persons with low measures will have a high probability of answering the easy items positively, and a low probability of answering the medium and hard items positively. Persons with medium measures will have a high probability of answering the easy and medium items positively, and a low probability of answering the hard items positively. Persons with high measures will have a high probability of answering the easy, medium and hard items positively. These conditions are tested through a Rasch analysis.
- c. Data were collected from persons on the items and scored dichotomously (0/1 or 1/2), as in, for example, but not limited to, wrong/right, no/yes, none/a lot, disagree/agree, some/often, bad/good, slow/fast.
- d. Each item is represented by a number, estimated from the data that represents its difficulty (called an **item parameter** in the mathematical representation of the Rasch Model) that does not vary for persons with different measures of the variable. Persons with different measures responding to the items have to agree on the difficulty of the items (such as easy, medium and hard, as used in this example). If the persons do not agree on an item difficulty, then this will be indicated by a poor fit to the measurement model, and then the item may be discarded as not belonging to a measure on this continuum.
- e. Each person is represented by a number, estimated from the data that represents his or her measure of the variable (called a **person parameter** in the mathematical representation of the Rasch Model) that does not vary for items of different difficulty along the continuum. If different items do not produce agreement on a person measure, then this will be indicated by a poor fit to the measurement model, and then one examines the person response pattern (and the items).
- f. Rasch measurement models use a probability function that allows for some variation in answering items such that, for example, a person with a high attitude measure may give a low response to an easy item, sometimes, or a person with a medium achievement measure might get a hard item right, sometimes. If the person response pattern shows too much disagreement with what is expected,

then it may be that the person has not answered the items properly or consistently, and that person's results may be discarded, or the item may be too hard or too easy, requiring it to be modified. In the mathematics of the model, the probability of answering correctly is related to the difference between the person measure and the item difficulty. In situations where there is a large positive difference between the person measure and item difficulty, then there is a strong probability of a correct response and, if there is a large negative difference, then there is a strong probability of an incorrect response. If the differences are not so large, the probabilities are changed appropriately (Waugh, 2007).

Equations for the Simple Logistic Model of Rasch

$$\begin{array}{l} \text{Probability of answering} \\ \text{positively (score 1)} \\ \text{for person n} \end{array} = \frac{e^{(B_n - \delta_i)}}{1 + e^{(B_n - \delta_i)}}$$

$$\begin{array}{l} \text{Probability of answering} \\ \text{negatively (score 0)} \\ \text{for person n} \end{array} = \frac{1}{1 + e^{(B_n - \delta_i)}}$$

Where:

e = natural logarithm base (e=2.7318)

B_n = parameter representing the measure (ability, attitude, performance) for person n

δ_i = parameter representing the difficulty for item i

Source: Rasch, 1960, 1980, 2010

These equations are solved from the data (entered in a text format) by taking logarithms and applying a conditional probability routine with a computer program such as RUMM (Rasch Unidimensional Measurement Models) (source: Andrich, 1988b; Rasch, 1960, 1980; Waugh, 2007).

The Extended Logistic Model of Rasch for Partial Credit Scoring

The Partial Credit Model of Rasch can be thought of as an extension of the Simple Logistic Model from two response categories to three or more response categories or outcomes (see Andrich, 1988b; Masters, 1988b; 1997). So the conditions, requirements and output of the Partial Credit Model are similar to the SLM, except that there are now more item parameters, and more item output and the equations are more complicated. The Partial Credit Model can be applied to any set of data scored, judged or answered in three or more ordered outcome categories where the level of outcome is conceptualised on a continuum from low to high.

Equations for the Partial Credit Model of Rasch

$$\begin{array}{l} \text{Probability of person } n \text{ scoring} \\ \text{in outcome category } x \text{ of item } i \\ \text{(for } x = 1,2,3,4 \dots M_i) \end{array} = \frac{e^{\sum_{j=1}^x (B_n - \delta_{ij})}}{1 + \sum_{k=1}^{M_i} e^{\sum_{j=1}^k (B_n - \delta_{ij})}}$$

$$\begin{array}{l} \text{Probability of person } n \text{ scoring} \\ \text{in outcome category } x \text{ of item } i \\ \text{(for } x = 0) \end{array} = \frac{1}{1 + \sum_{k=1}^{M_i} e^{\sum_{j=1}^k (B_n - \delta_{ij})}}$$

Where:

e = natural logarithm base ($e=2.7318$)

$\sum (B_n - \delta_{ij})$ is the sum of $B_n - \delta_{ij}$

B_n = a parameter representing the measure (ability, attitude, skill or performance) for person n
 $\delta_{i1}, \delta_{i2}, \delta_{i3}, \dots, \delta_{iM_i}$ = are a set of parameters for item i which jointly locate the model probability curves for item i . There are M_i item parameters for an item with $M_i + 1$ outcome categories.

(Source: Masters, 1997, p.859)

Rasch Data Analysis with the RUMM (2030) Computer Program

There are eight data analysis tests (output) provided in the creation of a linear, unidimensional scale by using the Partial Credit Model of Rasch with the RUMM 2030 computer program. This output is similar for the Partial Credit Model of Rasch and the SLM of Rasch. The only exception is that for the Simple Logistic Model there are no ordered thresholds, only one threshold (*RUMM 2030 Manual*, 2009; Waugh, 2007).

a. Testing that the response category is answered consistently and logically

The RUMM program does this with two outputs: one, it calculates threshold values between the response categories for each item (where there are odds of 1:1 of answering in adjacent categories) and, two, it provides response category curves showing the graphical relationship between the linear measure and the probability of answering each response category.

b. Testing for dimensionality

An item-trait test-of-fit is calculated as a chi-square with a corresponding probability of fit. It tests the interaction between the responses to the items and the person measures along the variable and shows the collective agreement for all items across persons of different measures along the scale. If there is no significant interaction, one can infer that a single parameter for each person can be used to accurately predict each person's response to all the different items along the scale (described by a single parameter for each item) and it is in this sense that we have a unidimensional measure (Andrich & van-Schoubroeck, 1989, pp. 479-481).

c. Testing for good global Item-Person Fit Statistics

The item-person test-of-fit examines the response patterns for items across persons and the person-item test-of-fit examines the response patterns for persons across items using residuals. Residuals are the differences between the actual responses and the expected responses as estimated from the parameters of the measurement model. When these residuals are summed and standardized, they will approximate a distribution with

a mean near zero and standard deviation near one, when the data fit a Rasch measurement model.

d. Person Separation Index

The RUMM program calculates a Person Separation Index that is constructed from a ratio of the estimated true variance among person measures and the estimated observed variance among person measures by using the estimates of the person measures and their standard errors. This tests whether the standard errors are much smaller than the differences between the person measures (Andrich & van-Schoubroeck, 1989, p. 483).

e. Testing for good individual item and person residuals

Residuals are the differences between the observed values and the expected values estimated from the parameters of the Rasch measurement model. It is instructive to examine these outputs as they give an indication of whether persons are answering items in a consistent way and they give an indication of individual person and individual item fit to the measurement model.

f. Item Characteristic Curves

Item Characteristic Curves examine how well the items differentiate between persons with measures above and below the item location. An Item Characteristic Curve also shows a comparison between the observed and expected proportions correct for a number of class intervals of persons.

g. Person Measure/Item Difficulty Map

The RUMM program produces two types of person measure/item difficulty maps. These maps show how the person measures are distributed along the variable and how the item difficulties are distributed along the same variable (measured in logits). They show which items are easy, which ones are of medium difficulty and which ones are hard. They show how well the item difficulties are targeted at the person measures. That is, they show whether the items are too easy or too hard for the persons being

measured and whether new items need to be added, or whether there are too many items of similar difficulty (some of which are thus not needed).

h. Testing for construct validity

If the items are conceptually ordered by increasing difficulty (downwards) and the perspectives are ordered by increasing difficulty (to the right), then this represents the structure behind the variable. In Rasch measurement, all the item difficulties are calculated on the same linear scale and so the item difficulties can be compared with their conceptualised order. In this case, the item difficulties increase vertically downwards for each perspective by item and they increase horizontally to the right for each item by perspective. This provides strong support for the structure of the variable as it was postulated before the data were collected and analysed (Note: taken from Waugh, 2003; 2005; 2007; 2010a; and 2010b).

Variables

The present study measured three variables: English Reading Comprehension, English Writing, and Attitude and Behaviour with regard to English learning.

Measuring English Reading Comprehension

In this study English Reading Comprehension consisted of 11 multiple-choice questions and two essay questions. The multiple-choice items were arranged in order of predicted difficulty and are presented in four sub-parts: 'Needs for Energy' (easy); 'Sunlight and Producers' (hard); 'Consumers and Decomposers' (harder); and 'Humans in the Food Chains' (harder). Under each sub-part, three questions were also ordered by difficulty. Low question numbers under each sub-part were designed to be less difficult than high question numbers under the same sub-part. Thus, for sub-part 'Needs for Energy', as an example, question number 1 was considered easy; question number 2 was considered hard; and question number 3 was considered harder. This design applied to the other three subparts. The reason was due to the expectation that the test was designed in such a way that the low question numbers were easier for the students to answer; the higher the question number, the more difficult the question was for students to answer. The prediction that a question is considered easier than another was based on Bloom's taxonomy levels of thinking and thus whether the students were expected to

find it easy or hard to understand and/or consume less time in finding the correct answer. The opposite was that when it was hard for the students to locate the correct answers from the text, or if it took some time for them to find the correct answer from the text or it needed a higher level of understanding, then the items were predicted to be harder.

Take sub-part ‘Need for Energy’ as an example. Question number 1 (‘What will happen if there is no food?’) is considered the least difficult of all, because the question is based on common knowledge. The students were expected to understand it easily as long as they understood the question. For question number 2 (‘From where do plants get energy?’), the students need to read the text carefully and to understand it well, even though the answer was clearly stated on the text. Therefore, it was considered harder than question number 1. Question number 3 (‘A food chain is...’) was also stated in the text but students needed to have a higher level of thinking (as in Bloom’s taxonomy) to understand the concept in order to be able to answer the question correctly. Thus, it was considered the hardest of all. Other sub-parts followed the same pattern. To avoid repetition, other sub-parts are not discussed here (refer to Appendix A for the complete English Reading test and Appendix B for English Reading Comprehension’s ordered scoring scheme). Items (correct answer and distracters) for each question have been ordered by difficulty as well, in terms of three degree of correctness: ‘Most Correct’ (Score 3), ‘Partly Correct or Partly Incorrect’ (Score 2), and ‘Least Correct’ (Score 1) (see Table 3.1).

Table 3.1
Item Difficulty Scoring Scheme for Reading Comprehension Test

Reading Comprehension Item		Least correct	Partly correct or partly incorrect	Most correct
		Score 1	Score 2	Score 3
Item difficulty				
Need for Energy				
1	What will happen if there is no food?	a	b	c
2	From where do plants get energy?	c	a	b
3	A food chain is	a	c	b
Sunlight and Producers				
4	What do plants use to grow?	c	a	b
5	What do we call plants?	a	b	c
6	Plants make	b	a	c
Consumers and Decomposers				
7	What are primary consumers?	b	c	a

8	What is an example of a carnivore?	b	a	c
9	Decomposers are	c	a	b
Humans in the Food Chain				
10	What do humans eat?	b	a	c
11	What do we call humans?	b	c	a
12	Is the Food Chain important? Why? Or Why not? Write your answer below.			
	Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.			

Source: Designed by Khairiah Syahabuddin for the present study (see Appendix A for the complete Reading Comprehension test)

There were two Essay questions given to the students on the Reading Comprehension test sheets. One question, which is question number 12, was related to the Reading Comprehension text, and another one, which is without numbering, was related to their overall Reading Comprehension learning. Question number 12 was marked as part of the Reading Comprehension test, but, the final question was asked to allow students to provide some further comments on learning English Reading Comprehension. The question was not marked for scaling and was not analysed as part of the Rasch measure. One student's answer from the Reading Comprehension test number 12 is shown in Appendix N.

Measuring English Writing Achievement

The English Writing Test consisted of two compulsory topics, which were 'My *Idul Fitri* Holiday' and 'My Family' (see Appendix C for English Writing test). Students were asked to write some paragraphs about the two topics. They were encouraged to write as much as they could within the assigned time of 70 minutes for both topics. They were allowed to choose to write either topic first. Both writing topics were marked using the scoring rubric which was designed by the researcher. The scoring rubric for English Writing involved categories conceptualised from easy to hard and harder still (see Table 3.2). The scoring rubric has three main sections: Paragraph(s) Organisation, Text Conventions, and Text Quality. Each of these has three levels of predicted difficulty from 'easy', to 'hard' to 'harder'. The reason behind the scaling difficulty of the content is that it is expected that measurement implies an ordering of items by difficulty. It is expected that Paragraph(s) Organisation is easier than Text Conventions and Text Quality, because Text Convention involves aspects within paragraphs. This is also true for Text Quality. It is predicted that students find

Text Quality the hardest of all, because it encompasses the other two (Text Conventions and Text Quality).

Paragraph(s) Organisation contains three sub-parts: 'Topic Sentence', 'Concluding Sentences', and 'Supporting Sentences, respectively. Text Conventions also have three levels of predicted difficulty from 'easy', to 'hard' and 'harder'. They contain 'Spelling', 'Punctuation/Capitalisation' and 'Grammar' respectively. Text Quality also has three levels of predicted difficulty from 'easy', to 'harder' and 'harder still'. It contains 'Readability', 'Style' (which includes Sentence fluency, such as varied length, good flow and rhythm, and varied structure), and 'Text Organisation', respectively. Each main section, which includes Paragraph(s) Organisation, Text Conventions, and Text Quality, has four categories of scoring: 'Excellent' for score 4, which is the highest score, 'Good' for score 3, 'Adequate' for score 2, and 'Poor' for score 1. The meaning of these is given in Table 3.2.

Level of difficulty for each of the three subparts follows the prediction of abilities needed by students to complete the tasks satisfactorily. For example, the Paragraph(s) organisation contains three subparts in order of difficulty: 'Topic Sentence' (easiest), 'Concluding Sentences' (hard), and 'Supporting Sentences' (harder). It is predicted that weaker-ability students find it easier to write a 'Topic Sentence', which might be regarded as simple knowledge, than to write a 'Concluding Sentence', and 'Supporting Sentences', which require a higher order of thinking (see the order of difficulty in Bloom's taxonomy from knowledge to application to analysis to evaluation). Similarly, students are expected to find writing 'Supporting Sentences, the hardest compared with the other two ('Topic Sentence' and 'Concluding Sentences'). The reason is weaker-ability students are expected to be able to write a topic sentence, but not necessarily be able to write a concluding sentence, which requires a higher order of thinking, as described in Bloom's taxonomy. Similarly, they might not produce some adequate and correct supporting sentences to support the main idea within the topic sentence. The Supporting Sentences were considered the hardest of all due to the expectation that the students needed more practice and knowledge to be able to write them than to write a topic sentence and a concluding sentence, and because they involve higher orders of thinking like analysis and synthesis.

Table 3.2
English Writing Scoring Rubric

		TEXT WRITING	EXCELLENT (Score 4)	GOOD (Score 3)	ADEQUATE (Score 2)	POOR (Score 1)
Paragraph(s) Organisation						
Easy	1	Main Topic	Main topic is clearly stated in a complete opening sentence with correct sentence structure.	Main topic is stated in an opening sentence but sentence structure is not correct AND/OR lacking detail.	Main topic is not clearly stated but sentence structure is correct.	Main topic is not clearly stated. Sentence structure is not correct.
Hard	2	Concluding Sentence	Concluding sentence restates topic with correct sentence structure.	Concluding sentence restates topic but sentence structure is not correct AND/OR lacking detail.	Concluding sentence is not clearly stated but sentence structure is correct.	Concluding sentence is not clearly stated. Sentence structure is not correct.
Harder	3	Supporting Sentences	Text has at least 3 detailed supporting sentences AND all sentences are on topic.	Text has at least 3 supporting sentences *Only 2 sentences are on topic. AND/OR *Sentences lack details	Text has at least 3 supporting sentences *Only 1 sentence is on topic. *Sentences lack details.	Text has supporting details BUT none are on topic.
Text Conventions						
Easy	4	Spelling	There are 0-2 spelling errors.	There are 3-4 spelling errors.	There are 5-6 spelling errors.	There are more than 6 spelling errors.
Hard	5	Punctuation/ Capitalisation	Text has 0-2 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 3-4 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 5-6 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has more than 6 errors in punctuation/ capitalisation, and noun-verb agreement.
Harder	6	Grammar	Text has 0-2 errors in noun-verb agreement.	Text has 3-4 errors in noun-verb agreement.	Text has 5-6 errors in noun-verb agreement.	Text has more than 6 errors in noun-verb agreement.
Text Quality						
Easy	7	Readability	Text is neat and readable with 0-2 marked out words or other corrections.	Text is neat and readable with 3-4 marked out words or other corrections.	Text is not neat and readable with 5-6 marked out words or other corrections.	Text is not neat and unreadable with numbers of marked out words or other corrections.
Hard	8	Style (Sentence fluency, e.g. varied length,	Text shows sentence fluency.	Text shows reasonable sentence fluency.	Text shows minimal sentence fluency.	Text lacking in sentence fluency.

		good flow rhythm, and varied structure)				
Harder	9	Text Enjoyability	Text is enjoyable to read	Text is quite enjoyable to read	Text is satisfactory to read	Text is not enjoyable to read

Source: Designed by Khairiah Syahabuddin for the present study (see Appendix C for the complete English writing test)

Questionnaire on Attitude and Behaviour with regard to Learning English

The Questionnaire on Attitude and Behaviour with regard to Learning English consisted of 18 statements and one essay question. The 18 statements are presented under seven sub-headings: Tasks for Listening, Tasks for Speaking, Tasks for Reading, Tasks for Writing, Student/student relationships, Student/teacher relationships and Common Views. Statements under each sub-heading have been ordered by difficulty, that is, easy, harder, and harder still. Take, for example, ‘Tasks for Listening’. It has three statements. Statement number 1 was considered easy; statement number 2 was considered hard; and statement number 3 was considered harder. Statements under the other sub-headings follow the same pattern. The reasoning behind this order relates to how hard it is expected for a student to hold each Attitude statement and how hard it is expected for a student to behave according to each Behaviour statement.

The arrangement of predicted items difficulties is in line with previous attitude/behaviour type Rasch measures performed by Waugh (Waugh, 2003; 2005; Waugh, 2010a; 2010b). Agreement indicated support for the structure of the variables. Each of the 18 statements was to be answered in two perspectives: ‘This is what I wish to happen’ (that is, Attitude) and ‘This is what does happen’ (that is, Behaviour), making an effective item total of 36. The Attitude statements under ‘This is what I wish to happen’ were considered easier than the Behaviour statements under ‘This is what does happen’ due to the expectation that to actually do ‘things’ at a high category level requires more effort than to merely think about what ought to be done ideally. Take for example statement number 1, ‘I pay attention to someone speaking English’. Under the Attitude perspective, one just thinks about how one ideally pays attention to someone speaking English; while under the Behaviour perspective, one has to make an effort to pay attention to someone speaking English and, because there are other distractions, one

has to focus and concentrate on the behaviour itself rather than just thinking about it. Under each perspective, the statements were to be answered in three ordered responses: ‘Most or all the time’ (score 3), ‘Some of the time’ (score 2), and ‘Never or rarely’ (Score 1). The single written-answer question at the end of the Attitude/Behaviour Questionnaire was designed to give an opportunity for students to express any other feelings or concerns regarding their attitude and behaviour with regard to learning English as a foreign language (see Appendix E for the full Attitude/Behaviour questionnaire). The essay question was not marked for scaling but was analysed qualitatively for common perspectives (details are explained in Chapter Eleven).

Table 3.3
Attitude and Behaviour Questionnaire

Questionnaire Item	This is what I wish to happen			This is what does happen		
	1	2	3	1	2	3
Item no.						
Tasks for Listening						
1-2 I pay attention to someone speaking English						
3-4 I ask others to speak slowly or repeat words in English						
5-6 I listen to English songs						
Tasks for Speaking						
7-8 I say new words several times in English						
9-10 I practise English with other students						
11-12 I start conversation in English with my friends						

Source: Designed by Khairiah Syahabuddin from several sources for this study. The full Attitude/Behaviour questionnaire is given in Appendix E.

The next chapter explains the Research Design for the present study.

CHAPTER FOUR

RESEARCH DESIGN

This chapter explains the design of the study. It covers the research strategy and design; mixed-method design; samples; pilot study; study and ethics approvals; control of extraneous variables in the quasi-experiments; test data collection, data entry, and data analysis of the quantitative and qualitative data.

Strategy and Design

This study involves a mixed-method design (Bergman, 2008; Clark & Creswell, 2008; Creswell & Clark, 2006) with two parts. The first part was a quasi-experiment with two intact groups involving a control group taught through monolingual instruction and an ‘experimental’ group taught through bilingual instruction, with pretest and posttest measures for English Writing, English Reading Comprehension, and Attitude/Behaviour with regard to Learning English. The data from both the pretests and posttests were analyzed using the computer program Rasch Unidimensional Measurement Models (RUMM2030) (Andrich, et al., 2010) in order to create linear measures. The linear measures were used with Mixed Design ANOVA, also known as Split-Plot ANOVA (with SPSS) with the pretest as the covariate, and Two-Way ANCOVA (with SPSS) (Pallant & Tennant, 2007).

The second part was the qualitative students’ written comments regarding their perceptions about learning English as a foreign language. The comments were taken from the same students who sat for the quantitative tests (N=394 Year 7 students taught English at bilingual schools and N=386 Year 7 students taught English at monolingual schools). The data were analysed by the analytic induction method to produce some general propositions or main issues (Miles & Huberman, 1994).

Mixed-Method Design

Mixed-method research is a design in which two different approaches are combined to collect both quantitative data and qualitative data, although the emphasis on each may not be the same (Bergman, 2009; Clark & Cresswell, 2008; McMurray et al., 2004). Bergman (2009) defined it as a design with a combination of at least one qualitative and at least one quantitative component in a single research project or program. More specifically, Morse and Niehaus (2009) defined it as a scientifically rigorous research project, driven by the inductive or deductive theoretical drive, and comprised of a qualitative or quantitative core component with qualitative-quantitative supplementary component(s). While the core component of the project is the complete method used to address the research question, the supplementary component is not. As it is a methodological strategy different from the method which is used to extend the investigation, the supplementary component is incomplete in itself or lacks some aspect of scientific rigour. It cannot stand alone and is regarded as complementary to the core component (Morse & Niehaus, 2009).

The mixed methods approach has three characteristic features: First, use of qualitative and quantitative approaches within a single research project; second, explicit focus on the link between approaches (triangulation); and third, emphasis on practical approaches to research problems (Denscombe, 2007). A mixed method design is believed to be stronger than one that uses only a single method, for example, one using only the quasi-experimental method (see for example, Bryman, 2006; Rocco et al., 2003; Collins et al., 2006; Greene et al., 1989). Descombe (2007) says that there are five reasons for this. The first is improved accuracy. Mixed methods design allows a researcher to use various methods for the same study, which improve the accuracy of the findings. Moreover, the design can assess the bias as well and it then becomes a valuable strategy for the development of research instruments. Greene et al. (1989) posit this use of a mixed methods approach can be used to “seek convergence, corroboration, correspondence of results from the different methods”. The second reason is that a mixed methods design gives a more complete picture of the answers to the research questions. The use of various designs with a mixture of data allows the researcher to have a description of matters under investigation. The third is compensating for strengths and weaknesses. Mixed methods design allows the researcher to be aware of

each method's strengths and weaknesses, which then give room to the researcher to recognise possible bias. By combining methods, it allows the researcher to compensate for the weakness. The fourth is developing the analysis. Mixed methods design allows the researcher to compare and contrast data better, which later can be used to better develop the analysis. The fifth is as an aid to sampling. The mixed method design allows the researcher to use information gathered as a basis to select respondents through a different and contrasting method (Denscombe, 2007).

Morse (1991) proposed two forms of methodology for mixed methods: (1) simultaneously, that is using both methods at the same time; and (2) sequentially, that is using the results of one method for planning the next method. She argues that using quantitative and qualitative methods to address the same research problem leads to the issue of weighing each method and their sequence in the study. The current study considers quantitative and qualitative methods simultaneously. Figure 4.1 illustrates the design of the study.

The present study took these comments into account and used a mixed-method approach involving a quasi-experimental design and the collection of some written comments from students at the posttest stage. The quasi-experimental design involved a control group (monolinguals) and an experimental group (bilinguals) with pretest and posttest measures. In order to gain some extra information students were asked to provide some written comments at the time of the posttest. These were analysed qualitatively and helped to provide a more complete answer to the research questions.

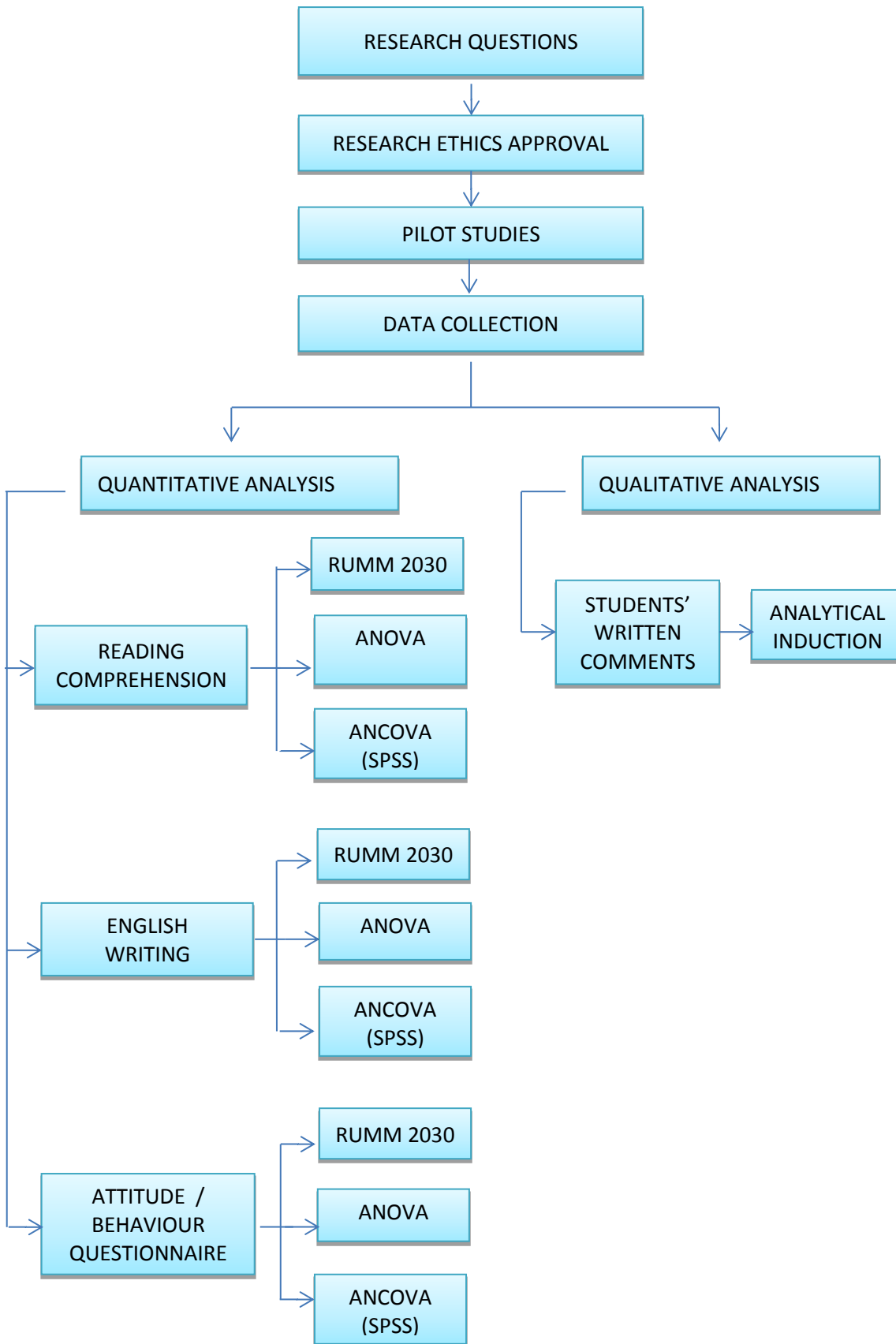


Figure 4.1 Design of the Study

Samples

Seven hundred and eighty male and female students who were enrolled in their first-year of middle schools (12/13 years old), consisting of 394 students from bilingual schools and 386 students from monolingual schools, were respondents for this study. The 780 students were taken from 13 State Middle Schools which are called 'SMPN'. 'SMPN' stands for '*Sekolah Menengah Pertama Negeri*', equivalent with Year 7 Primary School in the Australian context regarding the students' age. The schools are numbered under districts or suburbs: five schools having combined English and Bahasa Indonesia as medium of instruction, i.e. SMPN 1 (N=76), SMPN 2 (N=74), SMPN 6 (N=152), SMPN 13 (N=31), and SMPN 19 (N=61); and 8 schools using Bahasa Indonesia as a medium of instruction, i.e. SMPN 3 (N=76), SMPN 4 (N=48), SMPN 7 (N=50), SMPN 10 (N=52), SMPN 11 (N=19), SMPN 14 (N=33), SMPN 17 (N=63), and SMPN 18 (N=45).

The schools were taken from the same district or from neighbouring districts within one city, Banda Aceh. This was believed to be important to ensure that the students shared similar characteristics in terms of English teaching and learning experience; whether in bilingually-taught or monolingually-taught programs. Matching Bilingually-taught and monolingually-taught schools were chosen, as far as possible, by size and socio-economic status, but this also depended on voluntary agreement to participate. Schools surveyed for bilingual students were not also chosen to survey for monolingual students. It was intended that under no circumstances would students from either group interact with each other leading to collaboration between students.

The two groups shared the same characteristic regarding English learning, that is learning English as a foreign language. The two groups had the same English curriculum, similar syllabi, and they were taught with similar English teaching methodology. The students' after-school activities were similar. The major difference between them was the medium of instruction. Bilinguals were taught English as a subject using combined English and Bahasa Indonesia, while monolinguals were taught English using Bahasa Indonesia.

Almost all the students in both groups were of Acehnese (people of Aceh) origin and all spoke Bahasa Indonesia with their parents, siblings, school-mates and teachers.

The majority of them also spoke Acehnese as their heritage language with their parents, siblings, extended family, school-mates, and teachers. Only a few of them spoke English, read English books, watched TV programs, or were strongly involved in conversations in English with their classmates, outside the classrooms. English was rarely used in the monolingual classrooms but it was used in the bilingual classrooms. The only trivial difference between the two groups was the time for English lessons at school. The amount of English used for bilinguals varied, depending on the school, with a range of 38 hours to 57 hours per semester. The 38 hours per semester was made up of two hours for each meeting in a week for 19 meetings in a semester, while the 57 hours was made up of three hours for one or two meetings in a week for 19 meetings in a semester. The amount of English used for monolinguals was fixed, 38 hours per semester, consisting of two hours for one meeting for 19 meetings in a semester. Regarding this slight difference in the teaching-learning hours, the posttest for both groups was administered upon the completion of 32 hours of teaching-learning. This was considered important so as to ensure that both groups of students had been taught English as a subject for the same amount of time prior to the posttests being conducted.

Thus the samples were:

1. English Reading Comprehension and English Writing tests.
2. Questionnaire. All students who sat for the linguistic tests, that is, the English Reading Comprehension and English Writing tests, were requested to answer questionnaires on their attitudes and behaviour with regard to learning English.
3. Student Written Comments. All students who sat for the questionnaire test were requested to answer the last written item of the questionnaire which was about their experience in regard to learning English as a foreign language.
4. The 780 students were requested to sit for English Reading Comprehension and English Writing Experimental and Control Groups. Students who were taught English bilingually were assigned to the Experimental Groups, and students who were taught English monolingually were assigned to the Control Groups.

The study involved experimental and control groups with pretests and posttest using three measures: (1) English Reading Comprehension, (2) English Text Writing, and (3) an Attitude and Behaviour Questionnaire. Both control and experimental groups

were assessed under similar conditions and situations which involved the same English lesson content, similar lesson times and lesson duration, same homework, and same text books, but left the nature of the bilingual and monolingual teaching program unchanged. The only difference between the two groups was that the one group experienced bilingual teaching and learning processes with a slight extra amount of learning in class, while the second group experienced monolingual teaching and learning processes.

Teachers from bilingual schools and monolingual schools were counselled and monitored over the two months of the experiment to ensure that the controls on conditions and situations were working as they should. The points to consider were whether some changes in English lesson content, lesson times and lesson duration, homework, and text books occurred. Even though teachers in the bilingual and monolingual schools were different, they both shared similar awareness regarding the points of control that the study needed. The two-month experiment time was taken due to practicality issues and content issues. Practicality issues were teachers' willingness to participate in the research should the amount of experiment time exceed two months. Study issues were students' exposure to English language outside of the bilingual and monolingual classrooms. Since the focus of the study was to see the effect of bilingual and monolingual programs on student learning, controlling for extraneous English language variables outside the classroom was important. This was done by continual monitoring and counselling of teachers.

Table 4.1 shows the student numbers by control and experimental groups and by gender in control and experimental groups.

Table 4.1
Student Numbers by Control/Experimental Groups and by Gender in Control/Experimental Groups

Type of program	N	Gender	
		Female	Male
Bilingual	394	229	165
Monolingual	386	202	184

Instruments for English Tests

There are two researcher-designed English tests: (1) one for Writing English and (2) the second for English Reading Comprehension. No suitable tests could be found in the literature in line with the abilities of the Year 7 Banda Aceh students and so researcher-designed ones were used and piloted.

There are English Reading Comprehension Tests such as the ACER Neale Reading Comprehension Test (Neale, 1999) but this is just too advanced for the Banda Aceh students, so a researcher-designed one was prepared. The researcher-designed test involved students reading a passage about Food Chains taken from various sources in Aceh and then answering questions designed to test their comprehension of the text. There were twelve multiple choice items and two short written questions. There were three items on Sunlight and Producers; three items on Consumers and Decomposers; and two items on Humans in the Food Chains. One short written question was on the topic of the reading; while another one was a general question, which was on whether or not the students favoured Reading Comprehension. A scoring rubric was prepared so that the scored data was ready for Rasch analysis. Students were allowed to do the test within 40 minutes during school-time in school classrooms. The test was double-marked for consistency by the researcher and a pilot study was carried out prior to the main testing. Details of this test have been explained in Chapter Three (see Appendix B for English Reading Comprehension's ordered scoring scheme).

The English Writing Test contained two topics on which students wrote a number of paragraphs. The topics are 'My *Idul Fitri* Holiday' and 'My Family'. A scoring rubric was prepared in which three scoring aspects were ordered from easy to hard so that the scored data were ready for Rasch analysis. The scoring aspects are paragraph organization (topic sentence, concluding sentence, supporting sentences), text conventions (spelling, punctuation, grammar), and text quality (readability, style, and text enjoyability). The detail of this test has been explained in Chapter Three (see Appendix D for English Writing's ordered scoring scheme).

Questionnaire on Student Attitude/Behaviour

The researcher-designed questionnaire was based on a previously used Attitude and Behaviour with regard to English Learning Questionnaire published by Waugh, Bowering and Chayarathee (2005). This questionnaire was designed with Rasch measurement in mind and data from Year 6 primary students in Thailand produced a very good fit to a Rasch measurement model. Data collected with the questionnaire had a Person Separation Index of $PSI=0.92$ and a Cronbach Alpha of $\alpha=0.93$, showing good reliability. That is, a good linear, unidimensional scale was created with data collected with this instrument in Thailand.

Some changes were made to be applicable to Banda Aceh students. Waugh, Bowering and Chayarathee's Attitude and Behaviour (2005) had five tasks: Tasks for Group Work, Tasks for Meaning, Tasks for Reading Comprehension, tasks for Student/Student Relationship, and Student/Teacher Relationship. The present study had ten themes (or main issues): (1) Tasks on Learning English as a foreign language, (2) Student-Student Relationships, (3) Student-Teacher Relationships; (4) Personal Views on the Benefits of Learning English, (5) Common Views on the Benefits of Learning English, (6) Confidence and Achievement in Learning English, (7) Learning English through Media, (8) Family Support in Learning English, (9) Obstacles in Learning English, and (10) Other Views. The present study had three tasks with the same task name used in the previous study, namely: Tasks for reading comprehension, Student/Student Relationships, and Student/Teacher Relationship. However, the statements were partially amended. An example was Tasks for Reading Comprehension. The statements for the tasks of the previous study were: "I like to solve the problems/puzzles in reading assignments"; "I can put the story into the correct order"; "I can complete cloze exercises with the correct words"; and "I can find the correct answers to the reading questions". The statements for the present study were: "I guess the meaning of the English words in the text"; "I read words carefully in English"; and "I can read English at home on my own".

After some modifications were made, the revised questionnaire consisted of 21 items, each answered in two perspectives, making an effective item sample of 42. The two perspectives were: (1) 'This is what I wish to happen' (an attitude) and (2) 'This is

what does happen' (a behaviour). The 21 items were set out in order of difficulty and consisted of six parts: three items under Tasks for Listening, three items under Tasks for Speaking, three items under Tasks for Reading, three items under Tasks for Writing, three items under Student/Student Relationships, three items under Student/Teacher Relationship, and three items under Common Views. The questionnaire was translated into Bahasa Indonesia prior to the administering of the survey and a pilot study was carried out prior to the main testing. A 40 minute period was allocated to the questionnaire. Details of this test have been explained in Chapter Three (see Appendix E for the Revised Attitude/Behaviour Questionnaire, that is, the Attitude/Behaviour Questionnaire for the present study). The written comments were back translated into English upon being presented into Chapter Eleven.

Instruments for students' written comments

For the students' written comments, students were asked a final question on the Attitude and Behaviour Questionnaire "Is there anything else you would like to add about your experiences in learning English?" which had been translated into Bahasa Indonesian. Thus, the answers to this question were provided in Bahasa Indonesian by the students. There were 1846 comment, all were in Bahasa Indonesian), written by the students from both types of schools: with 1110 written comments from bilinguals, and 732 written comments from monolinguals. All the written comments were grouped into 13 themes or concepts and carefully tallied, as the following: Tasks for Listening, Tasks for Speaking, Tasks for Reading, Tasks for Writing, Student/student relationship, Student/teacher relationship, Personal Views on Benefit of Learning English as a Foreign Language, Common Views on Benefit of Learning English as a Foreign Language, Students' Confidence and Achievement, Learning English through Media, Family Support in Learning English as a Foreign Language, Obstacles of Learning English as a Foreign Language, and Other Views. Some comments were stated as students' quotations.

Administration of the Instruments

Procedure

First, the students from both schools where English was bilingually-taught and where English was monolingually-taught answered, as pretests, a 12 item multiple-

choice reading comprehension test and two topics on which to write, and a separate attitude and behaviour questionnaire. Second, after a two-month time which was equal to 16 teaching-learning meetings, the same students sat for the identical tests, as posttests.

One research assistant was present on the days of the pretest and posttest administration at each classroom to ensure the student-respondents understood the instructions. The classroom teachers were allowed to be present on the days of the tests but were not allowed to interfere in the test process in order to maintain the validity of the students' answers. The language of communication was the student-respondents' native language (Bahasa Indonesian). The test took about 40 minutes to answer. After two months, the same students sat for an identical test, as a posttest, under the same conditions.

The written comments were taken from the posttest Questionnaire, not from the pretest Questionnaire, because it was assumed that the students could give answers better after they had studied English longer, in this case, after a two-month period. Out of the 780 students (bilinguals, N=394; monolinguals, N=386) who completed the Attitude/Behaviour Pretest, 702 students (90%) provided an additional written comment. There were no problems encountered in gathering these data.

All the students' written comments, which were written in Bahasa Indonesia, were carefully read, translated into English, back-translated to ensure the accuracy of the translation, and classified into ten themes. Some student comments contained a number of themes. Each theme of a student's written comment was entered into a Microsoft Word table in the form of tallies which then were counted and checked again to ensure that they were accurately recorded.

Pilot Studies

A pilot study was carried out in October and November 2010 in Aceh, Indonesia, regarding the English Achievement Tests, that is the English Reading Comprehension, the English Writing Test, and the Attitude and Behaviour Questionnaire, on two types of students: bilinguals (N=31) and monolinguals (N= 28). The students who participated in the pilot study were excluded from participating in the main study that ran from January 2011 until April 2011. Some changes were made as a

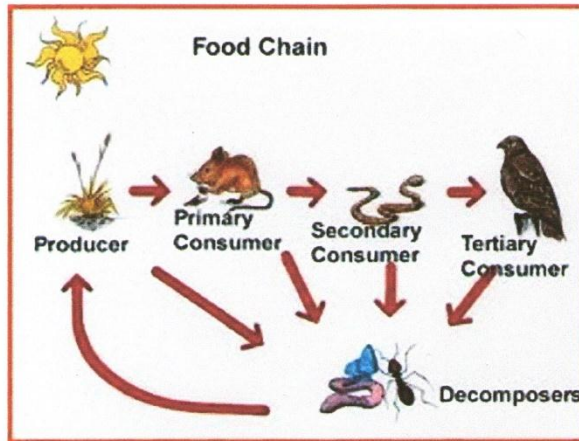
result of the pilot study for three variables: English Reading Comprehension, English Writing, and Attitude and Behaviour.

Pilot Study on English Reading Comprehension

A pilot study was carried out regarding this test. There were three changes resulting from the pilot study. First, the content and wording were changed, although the title remained unchanged. It appeared that the majority of the students found the text hard to understand. It is assumed that this was due to the text containing both complex and compound sentences and unfamiliar vocabulary. Therefore, it was changed to text containing a small number of complex and compound sentences and to more familiar vocabulary. For example the sentences like, '*Plants are called producers because they have the ability to convert the Sun's energy into chemicals that can be used to power life processes*' (Lines 4-5 of the original Reading Comprehension text) (see Appendix F for the original Reading Comprehension Text) were found to be difficult, since they contained a compound sentence (word 'and'), and complex sentence (word 'that'). The sentence of the text was then changed to '*All food chains begin with the sunlight. Plants use the sunlight to grow*'. The sentences were short, with repetitive words (*sunlight*) (see Appendix A for the revised English Reading Comprehension test, which was the English Comprehension test for the main data collection for English Reading Comprehension).

Second, a change was made to the pictures. Originally without pictures, the revised Reading Comprehension test contained colourful pictures. Pictures made the context clearer and more understandable for the students, especially for students with limited English in their initial year learning English as a foreign language. The following is an example of change in term of pictures. Figure 4.2 is the original Reading Comprehension Test Front Page, and Figure 4.1 is the revised Reading Comprehension Test Front Page.

FOOD CHAINS



Plants are important to our planet because they use photosynthesis to provide animal life with the energy they need to grow and to carry out life processes. When plants are eaten, this energy is transferred to another species. We call this a food chain.

There are three basic levels of a food chain.

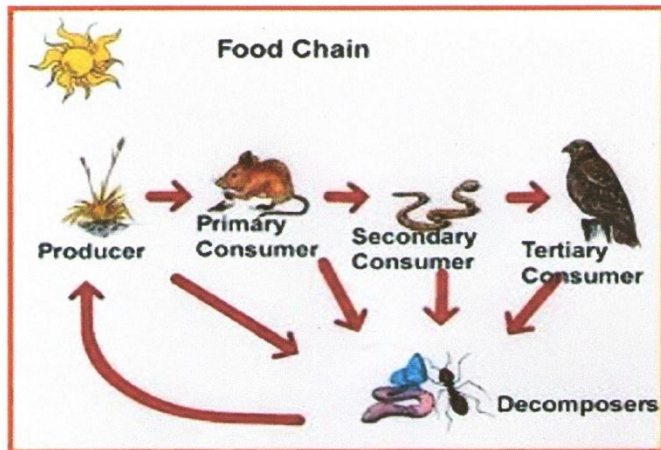
Level 1—Plants are called producers because they have the ability to convert the Sun's energy into chemicals that can be used to power life processes. Plants capture about 0.01 percent of the energy of the Sun's rays every day and are the basis of the whole food chain.

Level 2—Animals on the second level of a food chain are called primary consumers. We call them herbivores because they eat only plants for their energy. For example, deer eat shrub leaves, rabbits eat carrots and worms eat left litter. Primary consumers need to eat a lot of plants to meet their daily energy requirement. They eat approximately 10 per cent of the available plants and gain about 0.001 per cent of the Sun's original energy.

Level 3—Animals on the third level are called secondary consumers. They eat primary consumers for their energy, for example, lions eat deer, foxes eat rabbits and birds eat worms. Secondary consumers are good predators, but there are fewer of them because their energy sources are limited. Secondary consumers usually eat meat and we call them carnivores (if secondary consumers eat mainly insects, we call them insectivores). Secondary consumers gain about 0.0001 per cent of the Sun's original energy.

Further levels of the food chain are possible, but the longer a food chain is, the less energy left for the species at the top of the chain.

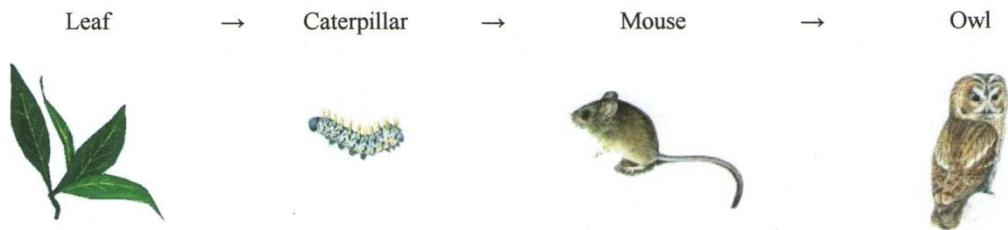
FOOD CHAINS



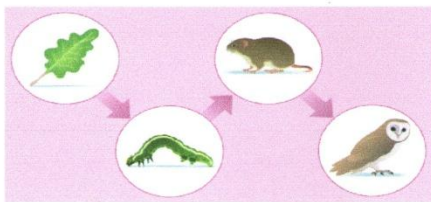
NEEDS FOR ENERGY

When you need energy in the morning before you come to school what do you do? Eat breakfast. Living things need food to give them energy. Plants get energy from sunlight. Caterpillars get their energy from plants. Mice get their energy from caterpillars. Owls get their energy from mice. Plants, caterpillars, mice and owls give energy to decomposers when die. The plants and animals in a food chain depend on each other to live. What would happen if all the plants died? The caterpillars would die. All the animals that depend on the caterpillars would die too. So every part of the chain is important.

Look at this picture representing a food chain using arrows,



Or this picture showing links or chains.



- The caterpillar eats a leaf
- The mouse eats the caterpillar
- The owl eats the mouse

The third change was related to time. The test was originally allocated 40 minutes but this was changed to 45 minutes. The original time was scheduled so that the Reading Comprehension Test and the Attitude and Behaviour Questionnaire were administered at one meeting of 80 minutes, consisting of 40 minutes for the Reading Comprehension and another 40 minutes for the Attitude and Behaviour Questionnaire. The Pilot Study suggested that students needed a little more time for the Reading Comprehension Test, while they needed less time for the Attitude and Behaviour Questionnaire. To accommodate the students' need, the allotted time was changed for the main tests to become 45 minutes for the Reading Comprehension test and 35 minutes for the Attitude and Behaviour Questionnaire.

Pilot Study on English Writing

Changes were made to the topics. The original topics were 'My Holiday' and 'My Hobby'. Due to the observations carried out during the Pilot Study and due to the English teachers' comments in Aceh, the topics for the main data collection were changed to 'My *Idul Fitri* Holiday' and 'My Family'. There were two reasons behind this. First, it appeared that the students doing the Pilot Study did not have adequate vocabulary to answer the original topics sufficiently well. The students were in their initial semester of learning English and the Aceh teachers believed that the existing vocabulary learnt prior to the tests did not cover the original topics of the test.

Second, the pilot study showed that the majority of the students did not know what to write on the selected topics. This might mean that the context of the topics did not suit their daily context. In Acehnese, as well as in the Indonesian context, people do not have a specific time for holiday and hobbies, unlike in some other countries, like Australia, where people prepare and arrange their holidays during school holiday times. Therefore, due to the above reasons, the topics for the main data collection were changed to topics that were believed to be better understood and suited to the Acehnese context. The revised first topic 'My *Idul Fitri* Holiday' is believed as a correct choice, since *Idul Fitri* is the Acehnese biggest religious celebration, which is similar to Christmas in the Australian context. It is the most enjoyed time of the year. The second topic 'My Family' is well-known to the students, since they live with their family. Therefore, it was expected that the revised topics could give better opportunities to

students to express their beliefs and feelings, than the original ones (see Appendix C for the Revised English Writing test, which was the English Writing Test for the main data collection).

Pilot Study on Attitude and Behaviour with regard to Learning English

A pilot study was carried out regarding this test. The changes made were in three parts: (1) wordings of two statements; and (2) an additional category with three additional questions. Firstly, it appeared that some wordings contained ambiguity for the students and, therefore, they were changed to avoid misunderstanding. The wording that were changed was under the category ‘Student/teacher relationship’ with the wordings originally “I like my English teacher teaching English reading”, revised to “I like the way my teacher teaches English Reading Comprehension”; and with the wording, originally, “I like my English teacher”, revised to “I like the way my teacher teaches English Writing”.

Secondly, answered essay questions provided by some students doing the pilot study of the Attitude and Behaviour questionnaire suggested to the author that further questions were needed for a better understanding of attitude and behaviour with regard to learning English. Therefore, a new category was added, called ‘Common Views’. It had three statements: “I like English because we use it in the classroom”, “I like English because it helps me in higher study”, and “I like English because it helps me go abroad”. The full, original Attitude and Behaviour questionnaire is given in Appendix G, and the full revised Attitude and Behaviour questionnaire, which was the Attitude/Behaviour test for the main data collection, is given in Appendix E.

Ethical Considerations and Study Approval

This research followed the strict regulations of the Edith Cowan University’s (ECU) ethical guidelines and the Human Research Ethic Committee.

Ethics Approvals

An ECU Ethics Form was completed and sent to the ECU Ethics Committee for approval in relation to this proposal. The ethical guidelines of ECU ensured that the research was conducted in a fair and acceptable manner. The following section

describes some ethical issues under the ECU ethical guidelines.

Informed consent letter

Prior to the conduct of research, the Head of Provincial Education of Aceh was approached and requested to give permission for the researcher to conduct research at all potential Middle Schools in Banda Aceh. The headmasters, regular English teachers, and students were approached and asked if they would be willing to participate in this research. Four different consent letters were sent to these different recipients. All consent letters which were given to all recipients for this study were translated into Bahasa Indonesia. The first was to the Head of Provincial Education of Aceh, requesting that the research data collection be approved in all potential state Middle Schools in Banda Aceh (see Appendix H for Request Letter to Conduct Research Data Collection to the Head of Provincial Education of Aceh). The second was to the Headmasters of the Middle Schools in Banda Aceh. The consent letters containing the explanation of the research to all potential headmasters were personally delivered upon approval from the Head of Provincial Education of Aceh (see Appendix I for Consent Letter to the Headmasters). The third was consent letters that were sent to regular English teachers of the Middle Schools (see Appendix J for Consent Letter to regular English teachers). The fourth was consent letters to the potential students of Middle Schools in Banda Aceh (see Appendix K for Consent Letter to students doing Reading Comprehension, Appendix L for English Writing, and Appendix M for Attitude/Behaviour Questionnaire). These consent letters were attached to the paper test and were collected on the day of each test. Headmasters' permission was given on the condition that the research data collection was conducted at their schools. The participation of the regular teachers depended on their willingness to let their students do the test at the normal English teaching-learning session time. Students' participation depended on their agreement to sit for three different tests for both pretest and posttest.

Possible risks to participants

There were no anticipated risks to participants in the study and all participation was voluntary. The statement in the letter outlined the purpose of the experiment and ensured the regular English teachers and students of confidentiality and anonymity, with the right to refuse to participate, and to withdraw from the experiment at any time. After

they had read the information and were satisfied, they were asked to sign a form of consent. Then, students signed the form of consent based on the conditions mentioned above, indicating their willingness to participate.

Payment for participation

There was no payment involved in the research. The participants agreed to participate in the research without any compensation. Certificates of appreciation and participation were provided to the regular English teachers.

Study Approval

The proposal for the research was presented at a post-graduate seminar at Edith Cowan University with two reviewers. There were minor changes made and then the proposal was submitted formally through the Faculty of Education and Arts for approval.

After ethics approval and study approval were granted, the data collection was conducted twice in Aceh, Indonesia: for the Pilot Study, that took about two months and for the main data collection, that took about four months.

Control of Extraneous Variables in the Quasi-Experiments

Ideally, the experimental process should be free from problems that were potentially contributed by people and the process of the experiment itself. However, problems can occur occasionally during experiments. In order to minimize problems of extraneous variables in the quasi-experiments, the researcher did the following:

1. She visited all potential state middle schools in Banda Aceh and approached headmasters and English teachers and informed them about the study. They were informed that the study required students who were learning EFL, as a subject, which was taught either bilingually or monolingually. They were well informed that the nature and characteristics of English as a medium of instruction were crucial. Should English be used minimally, then the schools were labeled as monolingually-taught. On the contrary, schools that used mostly English or at least 50% English as a

medium in teaching English subjects were labelled as bilingually-taught. This point was made clear so as to avoid respondent recruitment and contamination of results.

2. The headmasters and English teachers clearly understood that the nature or characteristics of their students' learning were important. Each type of student to be recruited as respondent should share a similar learning nature and characteristics. Nature and characteristics here refers to how they learnt English in the classroom and at school. Did they practise English at school and at home? How well did they do that? Did they read English books/magazines/novels/comics which were written in English? Did they watch English movies? Did they listen to English music? Those questions were posed in order to ensure that both types of students shared the same nature and characteristics in learning English. The answers from the headmasters and the English teachers were confirmed as that both types of students shared a similarity of learning nature and characteristics. Should the nature and characteristics on the way students' learning English be obviously different, they could not be taken as respondents. This study required that each type of student shared the same learning nature and characteristics, so that the main difference was the use of English as the medium of instruction *per se*. Students who were taught English using at least 50% English as the medium of instruction would fall into 'bilinguals'; while students who were taught using Bahasa Indonesia or Acehnese with at the most 10% English as the medium of instruction, would fall into 'monolinguals'. Both the most engaged English students and the least engaged English students were part of the study. The most engaged English students were labelled 'bilinguals' and the least engaged English students were labelled 'monolinguals'.
3. English teachers agreed to monitor the students' learning nature and characteristics and to report to the researcher should any changes in the students' learning nature and characteristics occur during the two-month experimental period. The focus of the study was to investigate the effect of the use of combined English and Bahasa Indonesian (hence called the bilingual program) and the majority use of Bahasa Indonesian to English (hence called the monolingual program) to the students' English improvement (English reading comprehension and English writing). Controlling for the extraneous English language variables outside the classroom was important. Teachers were asked to report to the researcher by notes taken during the

experimental process any deviation from the controls. From day one until the last day of the experimental process, there were no changes reported.

4. The researcher kept monitoring and counselling the English teachers during the experimental process.

Test Data Collection, and Data Entry

Test Data Collection

Data for English Reading Comprehension, English Writing, and Attitude and Behaviour were each collected from 780 students from 13 state Middle Schools in Banda Aceh. Students in both bilingual and monolingual classes completed a comprehension test, a writing test, and an Attitude and Behaviour questionnaire for the Pretest. After two months they completed another identical Writing Test, Comprehension Test and Attitude and Behaviour questionnaire for the posttest. The tests were conducted during the English lesson teaching time. It meant that the time for the test was not arranged out of that particular class's English lesson. The test of Reading Comprehension and the Attitude and Behaviour questionnaire were taken at one meeting-period which was about 90 minutes: 45 minutes for the Reading Comprehension test, and another 35 minutes for the Attitude and Behaviour. The test of Writing was conducted on the following day. Example answers from student tests are given in Appendix N (for English Reading Comprehension), Appendix O (for English Writing), and Appendix P (for Attitude/Behaviour Questionnaire)

Data for the students' written comments (qualitative part of the study) were taken from part of the Attitude and Behaviour Questionnaire that was the final question which was required for respondents to answer. The written answer for this final question was done during the time for the Attitude and Behaviour test time. There was no chance for students to cheat, and the written comments were truly done by the students themselves, so the data maintained its accuracy. All data for the three tests were taken home for data entry and analysis straight after the test was done. There were no problems with collecting these data.

Data Entry

Test data collected were sorted. Only data from students who sat for both the pretest and posttest for the three variables were taken for the study. Students who sat only for one test (either pretest only or posttest only) were withdrawn from the study. Students who sat for both pretest and posttest for only one variable, for example, Reading Comprehension only, but did not sit for English Writing and Attitude and Behaviour, were withdrawn from the study. The data from the correct respondents were then carefully numbered for student codes, marked, and entered to Excel files. Checks and rechecks were conducted before they were satisfactorily ready for analysis to ensure that the data were entered accurately.

Qualitative data collected for the students' written comments were carefully read, and tallied into Microsoft Word for analysis, and checked for accuracy again.

Data Analysis

Analysis of Reading Comprehension and Essay Writing Test data

The data from both pretests and posttests were analyzed using the computer program Rasch Unidimensional Measurement Model (RUMM2030) (Andrich, et al., 2010) to create linear scales. This computer program provides various statistics that fit the measurement model and various graphs relating to the linear scale (item difficulties with standard errors, person measures with standard errors, item response category curves, item characteristic curves for discrimination and targeting graphs by context variables like gender). The RUMM computer program was used to create linear scales for use in identifying students' linguistic achievements for both bilingually-taught school students and monolingually-taught school students. The measures obtained from the RUMM computer program were then used in the Mixed Design ANOVA, also known as Split-Plot ANOVA (with SPSS); and the Two-Way ANCOVA (with SPSS) (Pallant & Tennant, 2007) in order to compare pretest and posttest linear measures by control and experiment groups. The posttest measures were taken as the anchor (i.e. the process of using anchor values, which is a preset logit value assigned to a particular test/measure to be used as a reference value for determining the measurements or calibrations of other tests/measures) so that both pretest and posttest measures were on the same posttest scale and could be validly compared.

The items of the Reading Comprehension Test were conceptualized in order of difficulty and the three responses to each item were also conceptualized in order of difficulty (see Appendix A for Reading Comprehension Test, and Appendix B for English Reading Comprehension's ordered scoring scheme). Since the Rasch analysis calculated the actual item difficulties on the same linear scale, it was possible to compare the conceptualized item difficulties and their measured item difficulties. This is a construct validity test of the structure of the variable.

An ordered scoring rubric was conceptualized for the Writing Test (see Appendix C for English Writing Test and Appendix D for English Writing's ordered Scoring rubric). It is based on the structure of a good text: Text Organization, Text Conventions, and Text Quality. The ordered scoring has criteria for each of the three parts: for excellent (score 4), good (score 3), adequate (score 2), and poor (score 1). Text Organization (topic sentence is easy; concluding sentence is hard; and supporting sentences are harder); Text Conventions (punctuation/capitalisation is easy; spelling is hard; and grammar is harder); and Text Quality (readability is easy; style is hard; and text organisation is harder).

Details of the Rasch data analysis for Reading Comprehension are given in Chapter Five; and the details of the data analysis for the experimental measurement of Reading Comprehension are given in Chapter Six.

Analysis of Attitude and Behaviour Questionnaire Data

The questionnaire data for students were used to identify students' attitude and behaviour with regard to the English language program at the schools with special reference to bilingual and monolingual schools. These data were analysed by using the computer program Rasch Unidimensional Measurement Model (RUMM2030) (Andrich, et al., 2010) to create a linear, unidimensional measure. The measures were used in the Mixed Design ANOVA and the Two-Way ANCOVA (with SPSS) (Pallant, 2007), involving pretest/posttest, and control/experimental group measures with SPSS.

The items for the Attitude and Behaviour Questionnaire have been conceptually ordered by difficulty under six sub-headings: (1) Tasks for Listening, (2) Tasks for Speaking, (3) Tasks for Reading, (4) Tasks for Writing, (5) Student/Student

Relationships, and (6) Student/Teacher Relationships. Each item was answered in two perspectives by the students: (1) ‘This is what I wish to happen’ (an attitude) and (2) ‘This is what does happen’ (a behaviour), with a set of ordered response categories: scored 1 for never or rarely, scored 2 for some of the time, and scored 3 for most or all of the time. This provides a conceptual structure of item difficulties that can be compared with the actual Rasch-created item difficulties on the same linear scale to test the structure of the variable (its construct validity), like a science experiment. Details from the Rasch Analysis of Attitude and Behaviour measurement are given in Chapter Nine; and the details from the data analysis of the experimental measurement of Attitude and Behaviour are given in Chapter Ten.

Analysis of Students’ Written Comments

The students’ written comments were carefully read, classified into themes, tallied, and analysed using the analytic induction method (Miles and Huberman, 1994). Ten themes (abstractions or propositions) related to attitudes and behaviour with regard to the English language program at the schools were identified from the student comments. Details from the data analysis of Students’ Written Comments are given in Chapter Eleven.

The next chapter presents data analysis and findings from the Rasch measurement of Reading Comprehension.

CHAPTER FIVE

DATA ANALYSIS (PART 1)

RASCH MEASUREMENT OF READING COMPREHENSION

This chapter shows the findings of the Rasch analysis of the Reading Comprehension posttest data.

The Reading Comprehension Test (N= 12 items) was designed to measure the students' English Reading Comprehension at the posttest stage of the experiment. The data from the posttest were analysed with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010). The Reading Comprehension Posttest was scored in two categories of 0 (for wrong) and 1 (for correct) for 10 items, namely: items 1, 2, 3, 4, 5, 7, 8, 9, 10, and 11. These items were analysed using the Simple Logistic Model of Rasch (Rasch, 2010). Items 6 and 12 were analysed by using the Extended Logistic Model of Rasch because they had three scoring categories, i.e. 0, 1, and 2 (see Andrich, 1988a).

The English Reading Comprehension test was not part of the school assessment and was conducted for the purpose of this study. The test items were designed conceptually to fit on a continuum from easy to hard, and to be of the right difficulty range for the Indonesian students and to have all items gender and type of program (bilingual and monolingual) neutral. The test items were designed so as not to promote certain gender or type of program, for example, the bilingual program over the monolingual program, or vice versa. The idea behind the fit-on-continuum design is that students' ability could be predicted. It assumes that students with low ability would correctly answer the easy questions but would find it hard to answer correctly the moderately difficult questions and the difficult questions. The students with moderate ability would correctly answer both the easy questions and moderately difficult questions, but not the difficult questions; while students with high ability would

correctly answer the easy questions, the moderately difficult questions and the difficult questions.

The idea was to choose a text passage with questions of the right difficulty range and, at the same time, choose a neutral reading passage that was not beyond the students' ability (for both bilinguals and monolinguals) in terms of understanding the passage, but not favouring either female or male students. The chosen text passage was about The Food Chain, which was familiar to the students from their previous classroom work. This kind of passage had been taught in the students' previous lessons from year 1 to year 6 in Bahasa Indonesian and was considered important because the students learnt English for 4-5 months for one semester, with about 30-40 teaching hours within those 4-5 months. The majority of the students had very little, or no English at all, prior to their first year study in middle school. In Indonesia, English is taught as part of the curriculum in the first year of middle school, with an average of 38 teaching hours (1 teaching hour = 45 minutes).

Rasch measurement shows what should be expected in response to items if measurement at the metric, or linear level, is to be achieved (Pallant & Tennant, 2007). There are a number of statistical and graphical outputs to be considered using this analysis (see RUMM Manual, Waugh, 2003, 2005, 2010a, 2010b). These include Standardized Fit Residuals; Item-Trait Interaction (dimensionality); Person Separation Index, Individual Item Fit, Threshold Values, Response Category Curves, Item Characteristic Curves, Targeting, Differential Item Functioning (DIF) by Gender, Mean Measures by Gender, Differential Item Functioning (DIF) by Type of English programs, and Item Threshold Values.

The overall fit to the Rasch Model was not ideal. This meant that there was not good agreement about the item difficulties along the scale. Nevertheless, some other measurement aspects were satisfactory.

Output Analysis

Standardised Fit Residuals

Standardised Fit Residuals show the response patterns for items across students, and the response patterns for students across items. This interaction establishes the

overall fit statistics that confirms whether the item estimations contribute meaningfully to the measurement of one construct and sums over all items for each person and over all persons for each item. For an ideal well-targeted measure, which is not too easy or too difficult, the fit residual location mean is about zero and its standard deviation is about one (see Table 5.1) In this case, the fit is not ideal, but may be considered satisfactory when all the other RUMM output is considered.

Table 5.1
Overall Fit Statistics Reading Comprehension Measure (N=780, I=12)

	ITEM-PERSON INTERACTION			
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-1.59	1.54	-0.48
Standard Deviation	1.01	2.82	1.12	1.04

Notes:

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model.

Principal component analysis of residuals

The Principal Components analysis of the residuals showed that the item residuals loaded on a number of factors. However, deleting the not-so-well-fitting items did not improve the overall fit to the Rasch Model. The eigenvalue of the first component was 1.50 and that of the second component was 1.40, both of which are close to the chance level, supporting the finding of a unidimensional measure (see <http://www.rasch.org/rmt/rmt191h.htm>)

Item-trait interaction (dimensionality)

Item-trait interaction examines the consistency of agreement between students and the item difficulties (see Table 5.2). It checks that all the students agree that a particular item is easy, or of medium difficulty, or hard. In this case, the item-trait chi-square showed less than ideal agreement amongst the students on the difficulties of the items along the scale. It would seem that, in many cases, half of the students with medium measures said the item was easy and half said it was of medium difficulty. A similar situation occurred for some of the students with low measures.

Table 5.2
Item-Trait Interaction for Reading Comprehension Scale

Total Item Chi-Square	281.83
Separation Index	0.55
Total Degree of Freedom	72.00
Total Chi-Square Probability	0.00

Notes:

1. The item-trait interaction test is a chi-square. The results indicate that there was less than ideal collective agreement amongst the students about the item difficulties along the scale.
2. All numbers are given to two decimal points because the errors are only up to two decimal points.

The item-trait interaction chi-square was $\chi^2 = 200.4$, $df = 60$ with $p=0.00$ and, while five out of the 12 items fitted the model, they did not fit ideally well. It may be that some guessing of item answers by students at different measures along the scale was responsible for this non-agreement of the item difficulties, perhaps because of their low English Reading Comprehension standard and because of their desire to perform well. It seemed that there could be more than one scale present (or a number of factors or dimensions), but other data did not support this.

Person Separation Index

The Person Separation Index is an estimate of the true score variance among the students and the estimated observed score variance using the estimates of their ability measures and the standard error of these measures (Andrich & van Schoubroeck, 1989). For a good measure, the index is desired to be 0.75 or greater. The Person Separation Index, constructed as the ratio of the estimated true variance among the persons and the estimated observed variance among the persons, using the estimates of their locations and the standard errors of those locations, was =0.55, and is low. The Cronbach Alpha (Cronbach, 1951), based on the raw scores, was also low at 0.68. This indicates that the measures were only moderately well separated in comparison to the errors. It appears that there was insufficient variation in student measures along the scale in spite of the large sample size and, in retrospect; it seems that the 12 items were too similar and perhaps too similarly focused on the same content.

Individual Item Fit

Table 5.3
Locations, Standard Errors, residuals and Chi-Squares for Reading Comprehension Items

Item	Location	SE	Residual	ChiSq	DF	Probability
1	-0.545	0.122	-2.028	9.896	5	0.08
2	-0.118	0.077	-4.010	22.588	5	0.00
3	-0.932	0.140	-2.431	9.536	5	0.09
4	-1.098	0.149	-1.106	5.106	5	0.40
5	-0.296	0.113	-2.115	7.232	5	0.20
6	1.330	0.061	1.379	43.693	5	0.00
7	1.244	0.083	2.581	14.083	5	0.02
8	-0.848	0.136	-3.043	14.309	5	0.01
9	-0.505	0.121	-2.512	14.902	5	0.01
10	-0.341	0.115	-2.050	11.346	5	0.04
11	-0.749	0.131	-2.945	16.363	5	0.01
12	2.860	0.088	-1.917	45.087	5	0.00

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.
3. df means degrees of freedom. Probability is based on the chi-square fit to the measurement model.

The RUMM 2030 program calculates individual item fit to the measurement model. The fit statistics for all items for Reading Comprehension measure are presented in Table 5.3. While five items fitted the model, the other seven do not discriminate very well as expected from the model.

Raw residuals are the differences between the observed and expected responses. Standardized fit residuals are the differences adjusted to their standard deviations and they should be within the range $-2 < x < +2$. Table 5.3 indicates that some residuals are not ideal and that some items have not produced data that have a good fit to the measurement model.

Threshold Values

Table 5.4

Item Thresholds-uncentralised (Item=12, Number=780) for Reading Comprehension Measure

Item	Item Location	Mean Threshold	Thresholds	
			1	2
1	-.06	-.06	-.064	
2	-.45	-.45	-.454	
3	-.99	-.99	-.995	
4	-.38	-.38	-.379	
5	-.32	-.32	-.323	
6	+.98	.98	+.137	+1.833
7	1.34	1.34	1.337	
8	-.88	-.88	-.880	
9	-.49	-.49	-.488	
10	-.57	-.57	-.567	
11	-.52	-.52	-.519	
12	2.35	2.35	2.277	+2.418

Thresholds are points between nearby categories where the odds are 1:1 of answering in either category. The thresholds for items 6 and 12, which have three response categories (items scored 1, 2 or 3) are ordered, meaning that the response categories were used consistently and logically (see Table 5.4).

Scoring Category Curve

Scoring Category Curves represent the relationship between the probabilities of scoring in each category and person measures along the scale. Some items have two levels of scoring: zero for an ‘incorrect’ answer and one for a ‘correct’ answer for each item, and two items have three levels of scoring: zero is for an ‘incorrect’, one is for ‘partly correct’ and two is for a ‘correct’ answer.

Figure 5.1 is the Scoring Category Curve for item 1 of Reading Comprehension, which was scored dichotomously. The figure of its Scoring Category Curve shows that the scoring was done consistently and logically. When students have a low measure on item 1, then they have a high probability of answering item 1 incorrectly, thus receiving a zero score. When they have a high measure on item 1, they have a high probability in answering item 1 correctly, thus receiving score 1. The scoring for another ten items was checked and they were satisfactory too.

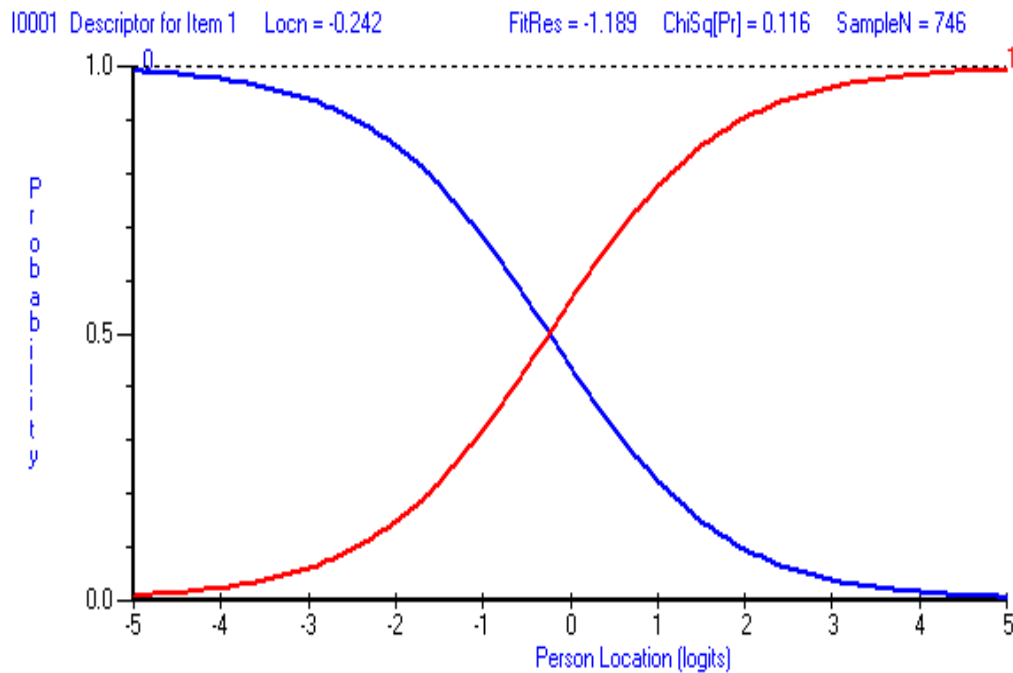


Figure 5.1 Item Category Curve for Item 1

Notes:

1. Blue colour curve shows the probability of scoring 0.
2. Red colour curve shows the probability of scoring 1.

Figure 5.2 is the Scoring Category Curve for item 6, scored 0, 1, 2. The figure of its Scoring Category Curve shows that the scoring was done consistently and logically. The Category 0 curve shows that if a Reading Comprehension Measure is located at -0.5 logits, the probability of answering the item correctly is around zero and, if the Reading Comprehension Measure is located at +5.0 logits, the probability of answering the item correctly is around 1. If a Reading Comprehension Measure is located at -5.0 logits, the probability of answering the item incorrectly is around 1. If the Reading Comprehension Measure is located at around +1 logits, the probability of answering the item partly correctly (and scoring 1) is around 0.5. If the Reading Comprehension measure is around +6.0 logits, then the probability of answering correctly is close to 1. The Category Scoring Curves for all the other 11 items were checked and they were satisfactory too.

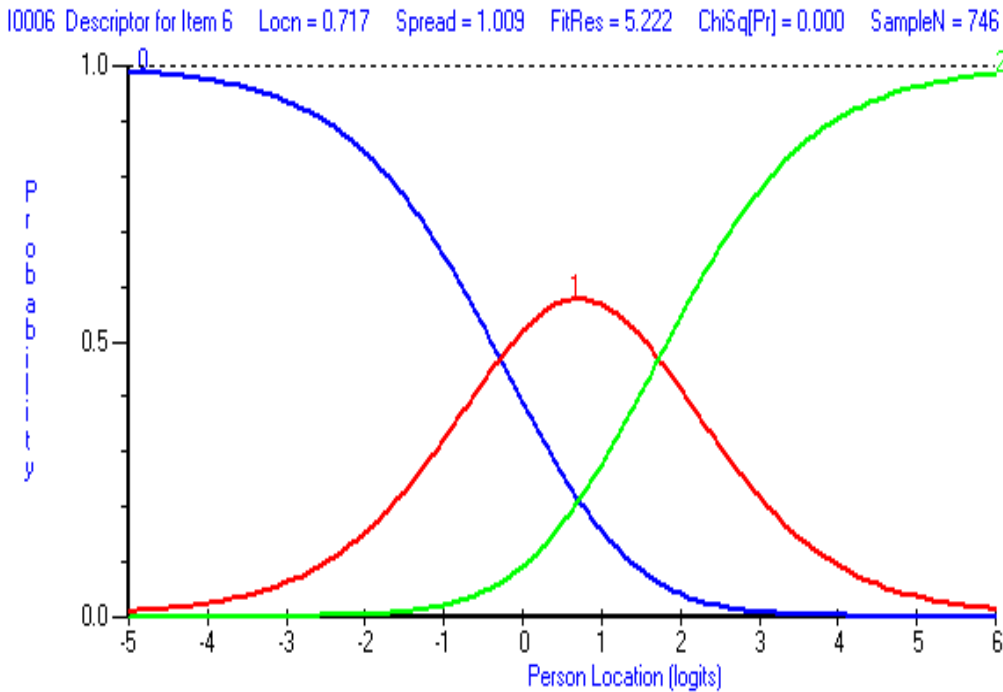


Figure 5.2 Item Category Scoring Curve for Item 6

Notes:

1. Blue colour curve shows the probability of scoring 0.
2. Red colour curve shows the probability of scoring 1.
3. Green colour curve shows the probability of scoring 3.

Item Characteristic Curves

Item Characteristic Curves show the relationship between the expected response score and the student measure, so that the researcher can examine how well the item differentiates between persons with measures above and below the item location. The observed means, shown as dots, in the seven class intervals are close to the ogive curve for item 1 (Figure 5.3) but not for item 6 (Figure 5.4) which shows that the item is not discriminating very well.

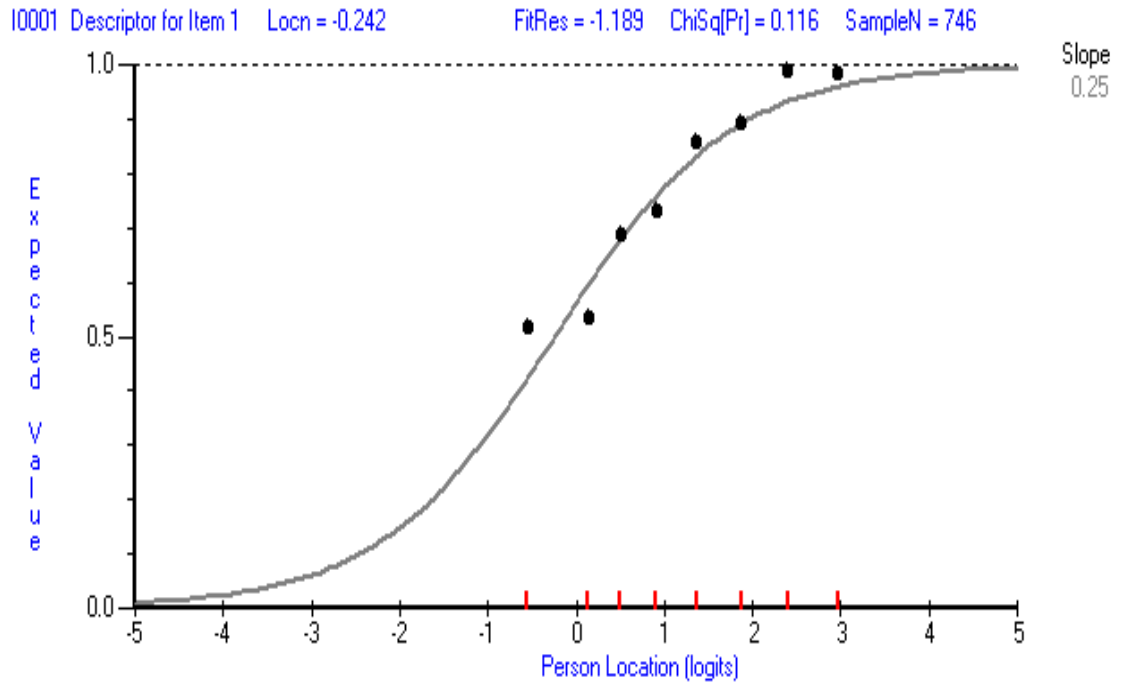


Figure 5.3 Item Characteristic Curve for Item 1

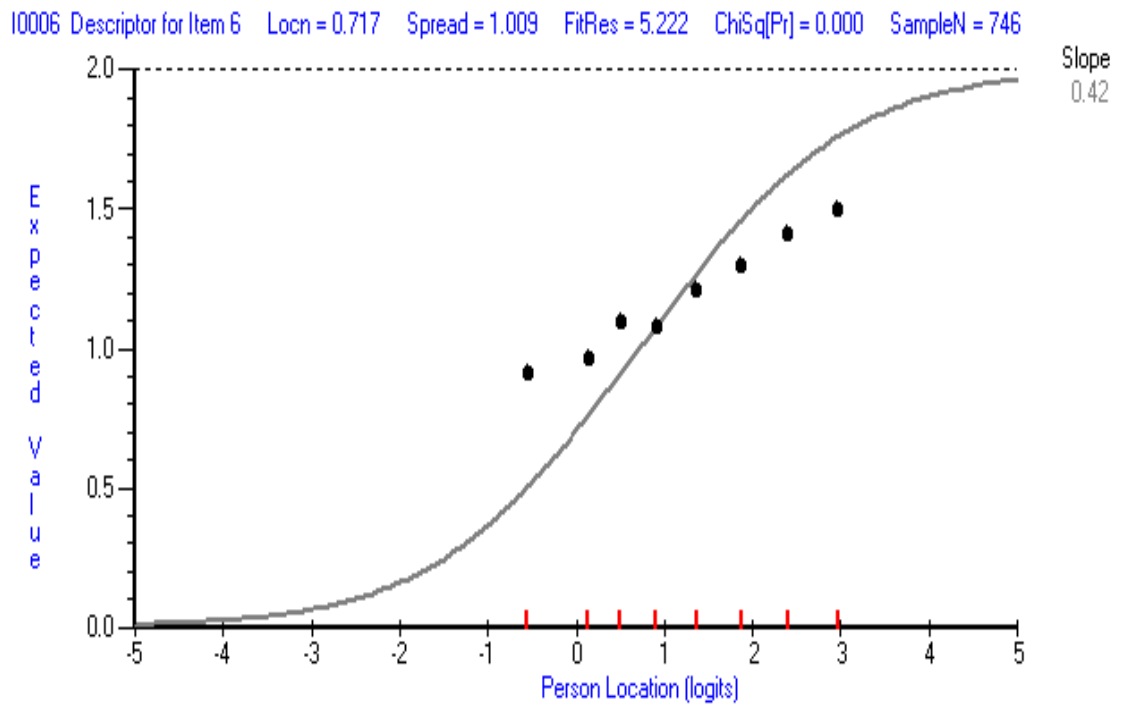


Figure 5.4 Item Characteristic Curve for Item 6

The Item Characteristic Curves were satisfactory in most cases without being ideal and targeting was reasonable – the student measures ranged from -0.8 to +3.9 logits and the item thresholds ranged from -1.2 to +2.6 logits.

Person-Item Threshold Distribution (Targeting)

Targeting shows how well the item difficulties are distributed along the scale against the student measures. A scale with a good measure should have a range of item difficulties that correspond to students with different student measures (Maley & Bond, 2011). Figure 5.5 shows student measures of Reading Comprehension from low to high on the top of the scale and item thresholds ranging from easy to hard on the bottom of the scale. The distribution graph showed that there were insufficient hard items to cater for the high measuring students and insufficient medium difficulty items to cater for the students with medium level measures. Before the scale is used in the future, more items should be added.

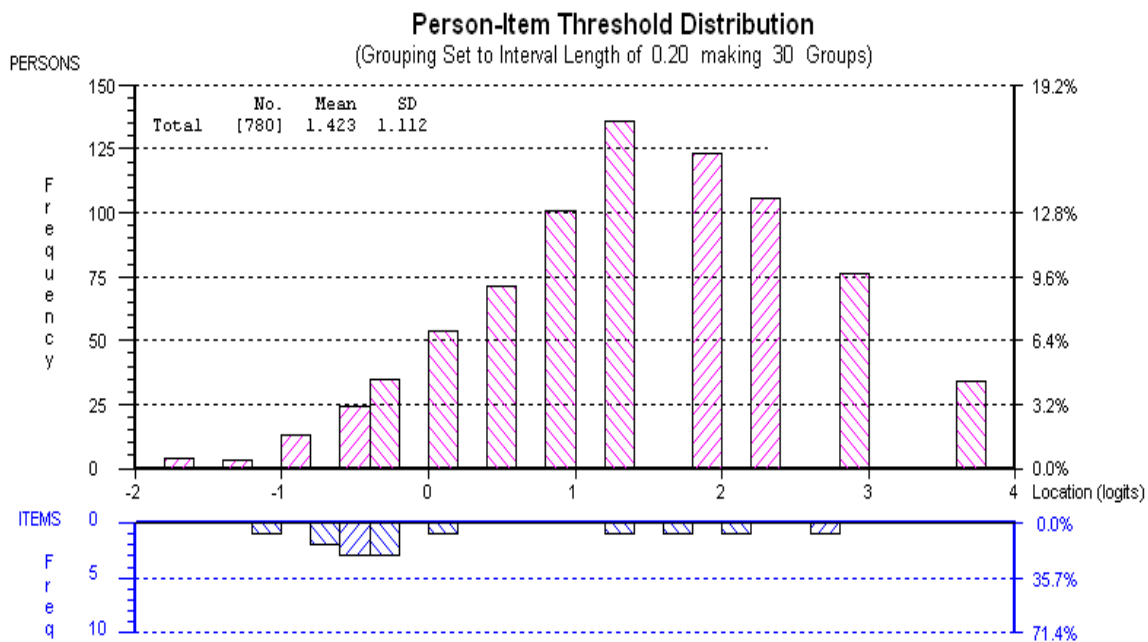


Figure 5.5 Person-Item Threshold Distribution for Reading Comprehension

The distribution of students and item difficulties are also illustrated in Figure 5.6, in which students’ reading comprehension and item ‘difficulties’ were calibrated on the same scale.

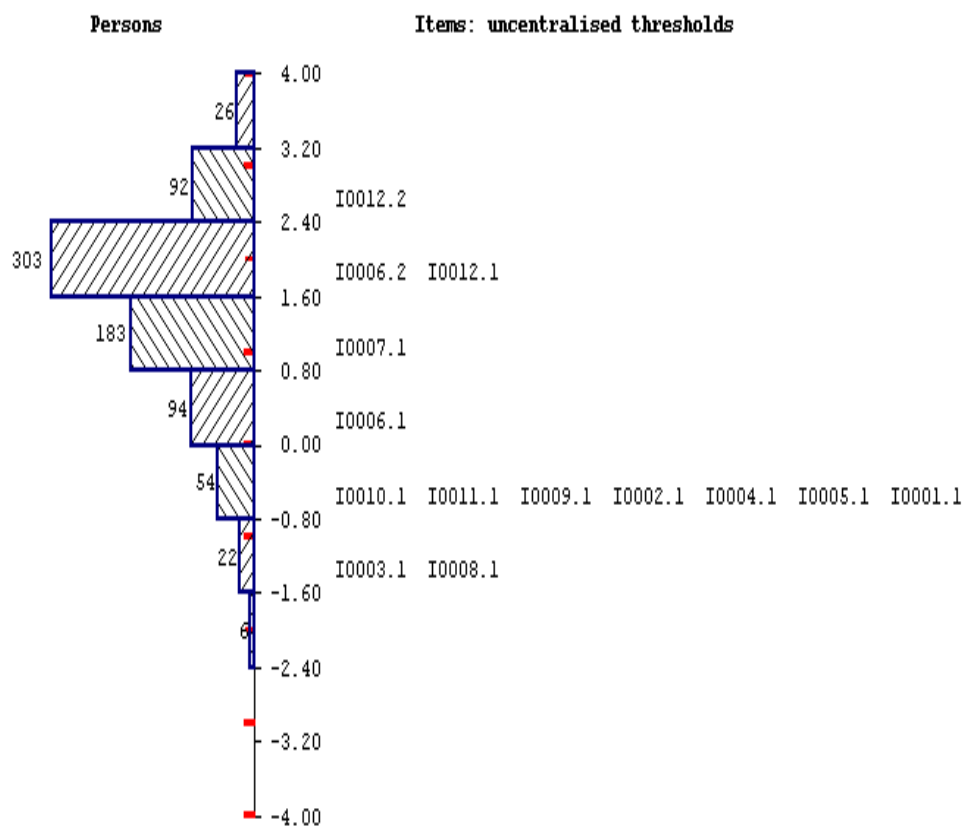


Figure 5.6 Students' Reading Comprehension and Item Thresholds Calibrated on the Same Scale

Notes:

1. Person measures are vertically left
2. Uncentralised thresholds are vertically right. I0012.2 means Item 12, threshold 2, and so on.

Differential Item Functioning (DIF) by Gender

Items 10 and 12 showed no statistically significant main effect for gender, with girls scoring higher ($F=2.56$, $df=6,1$, $p=0.11$) and ($F=5.26$, $df=6,1$, $p=0.02$) respectively, in which, the critical level is $p<0.01$. Item 10 was concerned with the relationship between humans and food in the Food Chain and Item 12 was concerned with from where the passage of writing on Food Chains was taken. While the items were not designed so that answers were item-dependent, it was likely in retrospect that some item dependency occurred because the item content was perhaps too similar. Item 10 and 12 graphs are shown in Figure 5.7 and Figure 5.8. Others, such as items 1, 3 and 6, also showed that girls performed better on the Reading Comprehension test than the boys, but these too were not statistically significant (see, for example, Figure 5.9).

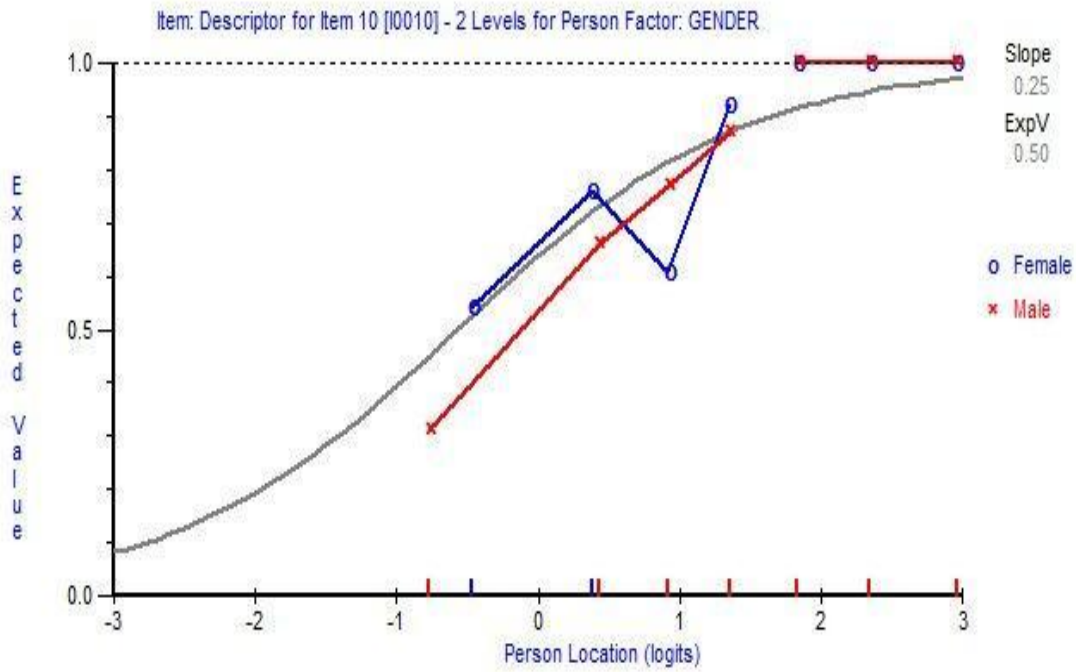


Figure 5.7 Item Characteristic Curves by Gender for Reading Comprehension Item 10

Notes.

1. Broken lines are due to technical errors of Item Characteristic Curves for Reading Comprehension Measure.
2. For the graph: No statistically significant main effect for females, $F=2.56$, $df=6,1$, $p=0.11$.

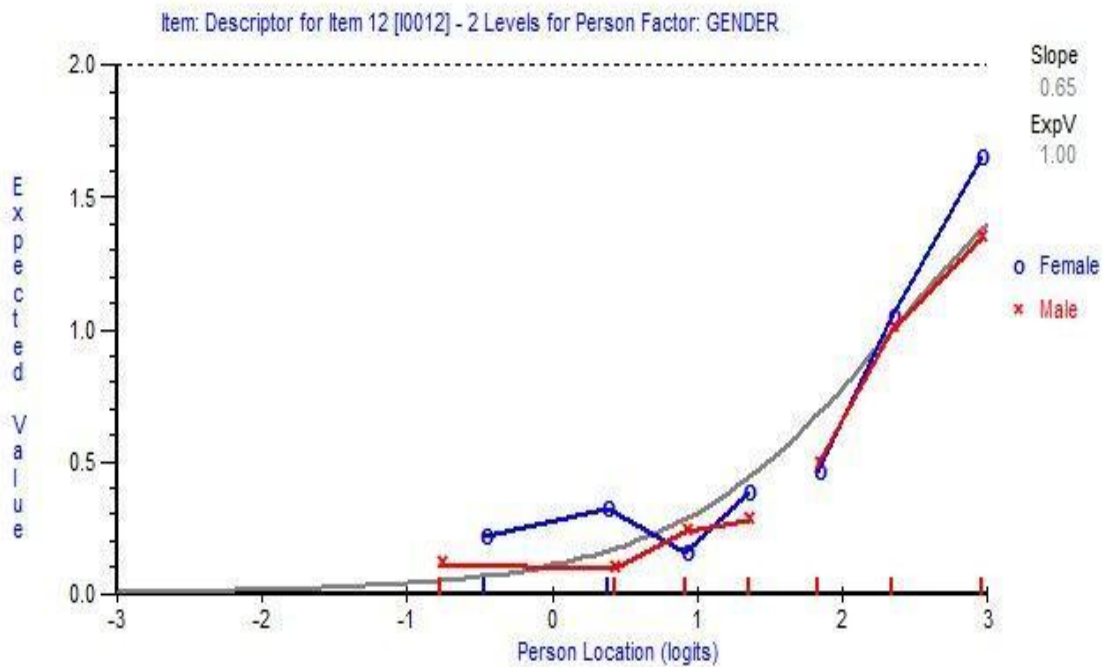


Figure 5.8 Item Characteristic Curves by Gender for Reading Comprehension Item 12

Note. No statistically significant main effect for females, $F=5.26$, $df=6,1$, $p=0.02$

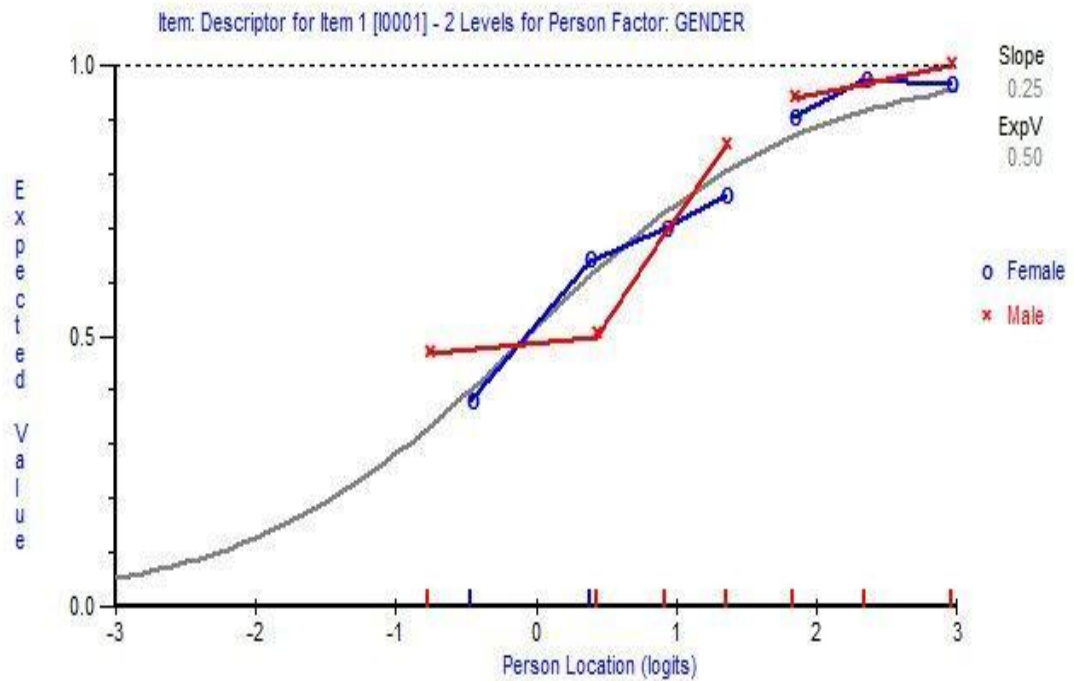


Figure 5.9 Item Characteristic Curves by Gender for Reading Comprehension Item 1

Note. No statistically significant interaction effect, $F=1.05$, $df=6,1$, $p=0.31$

Differential Item Functioning (DIF) by Type of Language Instruction

Items 2, 5, 6, and 9 showed a statistically significant main effect for type of teaching (bilingual or monolingual). For example with item 2, $F=15.37$, $df=6,1$, and $p=0.00009$, with bilingual teaching better; and with item 6, $F=43.45$, $df=5,1$ and $p=0.00000$ with bilingual teaching better. Figure 5.10 shows the Item Characteristic Curves by Type of Language Program (Bilingual/Monolingual) for Reading Comprehension item 2. Other items, such as item 8, showed that bilinguals did better than monolinguals, (see Figure 5.11), but this was not statistically significantly higher ($F=1.17$, $df=6,1$, $p=0.09$).

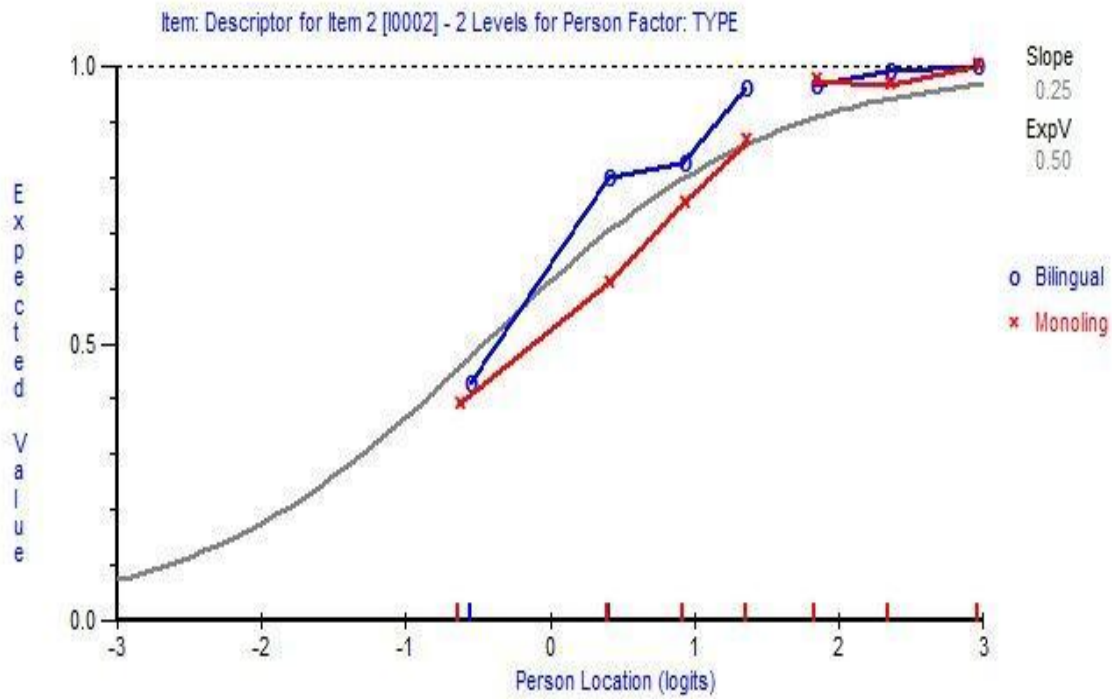


Figure 5.10 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 2

Note. There is a main effect, $F=15.37$, $df=6,1$, $p=0.00009$

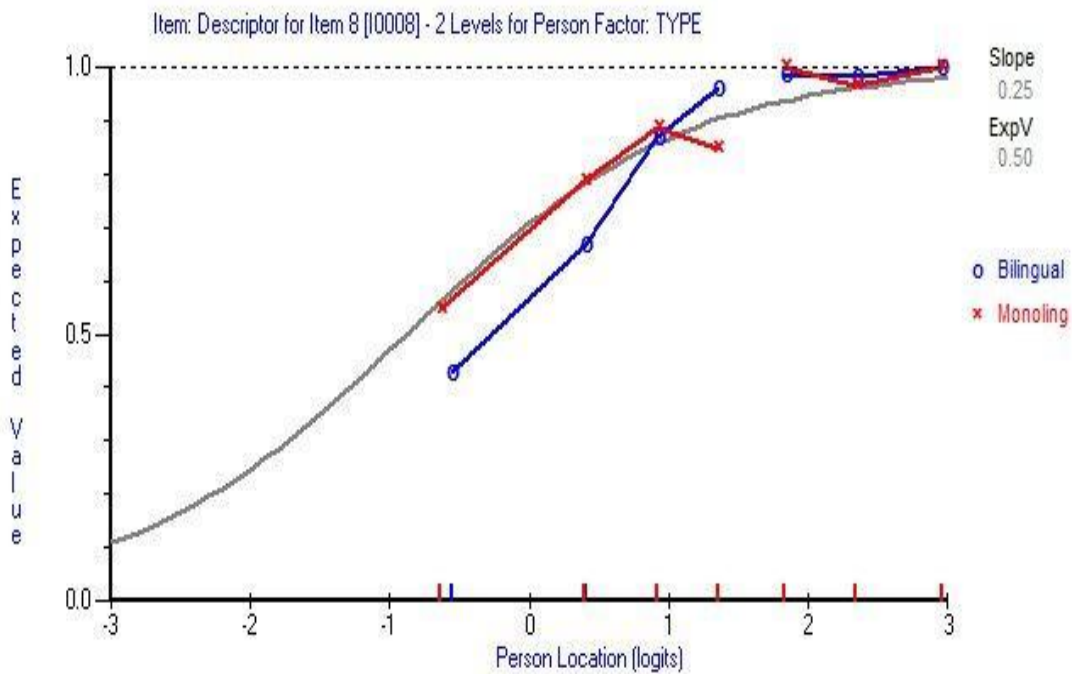


Figure 5.11 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 8.

Note. No statistically significant main effect, $F=2.90$, $df=6,1$, $p=0.09$

Targeting by Gender and Type of Language Instruction

Overall, girls performed statistically significantly better in the English Reading Comprehension than boys (mean 1.636 logits with SD = 1.05, and mean 1.430 logits with SD = 1.18, respectively) and $F=6.62$, $df=1,778$ and $p=0.01$ (see Figure 5.12). In addition, bilinguals performed statistically significantly better in the English Reading Comprehension than monolinguals (mean 2.094 logits with SD = 0.85, and mean = 0.982 logits with SD = 1.07 respectively) and $F=257$, $df=778,1$ and $p=0.00000$ (see Figure 5.13).

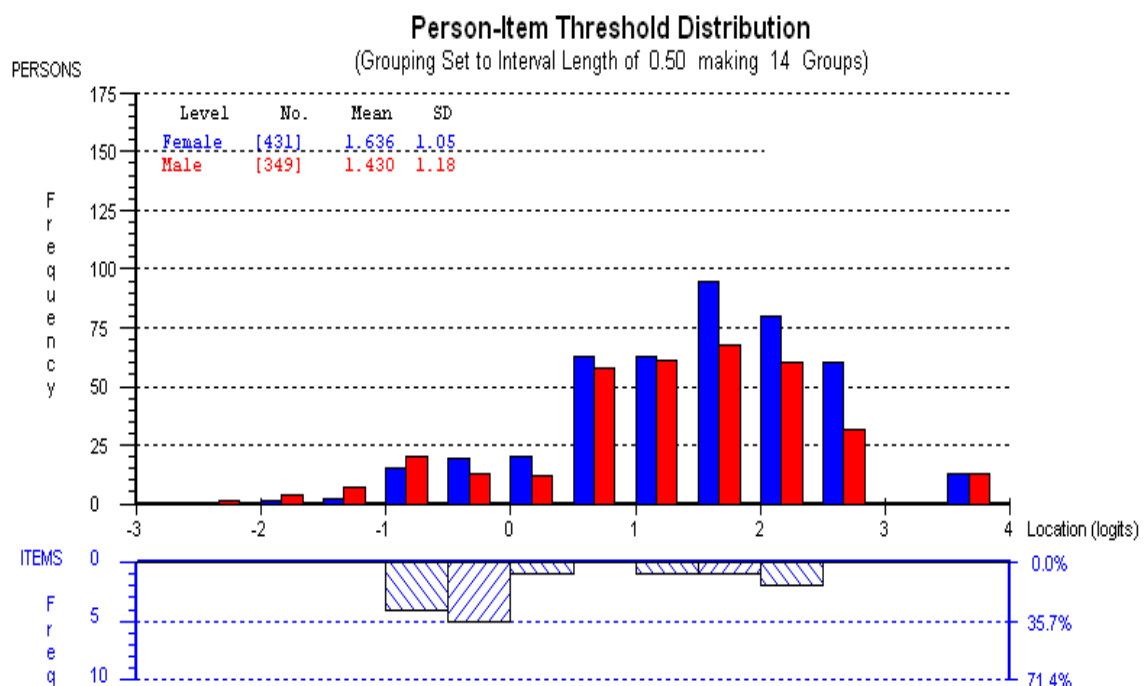


Figure 5.12 Targeting of Reading Comprehension by Gender

Notes:

1. The person measures are on the top-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the bottom-side from easy (LHS) to hard (RHS). $F= 6.62$, $df=1,778$, $p=0.01$, which is statistically significant for females.

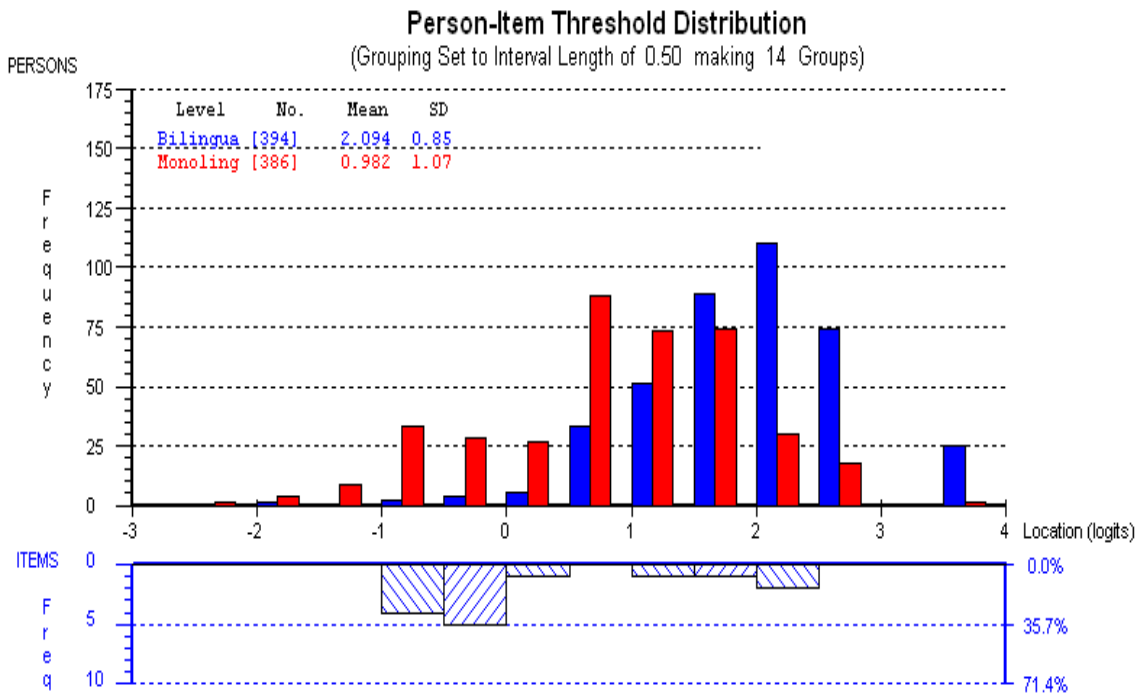


Figure 5.13 Targeting of Reading Comprehension by Type of Language Instruction

Note: $F= 257.89, df=1,778, p=0.00000$, which is statistically significant for bilingual teaching

Scale of Item Difficulties

Table 5.5
Order of Difficulty of Items on the Linear Scale

<i>Item No.</i>	<i>Item Location</i>	<i>Item Statements on the Food Chain</i>	<i>About</i>
3 (easy)	-0.995	Need for energy	link between plants and animals
8	-0.880	Consumers and decomposers	example of carnivore
10	-0.567	Humans in the food chain	what do humans eat?
11	-0.519	Humans in the food chain	what do we call humans?
9	-0.488	Consumers and decomposers	what decomposers eat?
2	-0.454	Need for energy	where do plants get their energy?
4	-0.379	Sunlight and producers	what do plants use to grow?
5	-0.323	Sunlight and producers	what do we call plants?
1	-0.064	Need for energy	what happens if there is no food?
6	0.985	Sunlight and producers	plant food
7	1.337	Consumers and decomposers	what are primary consumers?
12 (hard)	2.348	Humans in the food chain	why is the food chain important?

The items were ordered from easy to hard on a linear scale (see Table 5.5) so that it can be seen which items are easy and which are hard. The easiest item involved the need for energy and the link between plants and animals, as expected (item 3

difficulty = -0.995 logits). The hardest item involved explaining why the food chain was important, as expected (item 12 difficulty = +2.348 logits).

Discussion and Summary

After review by some teachers and modification, the passage of Reading Comprehension was initially considered appropriate in length for the second language students as again judged by their teachers in Aceh, and the topic was known to the students through previous teaching. Some teachers in Aceh reviewed the Reading Comprehension passage (and the items) and they suggested decreasing the difficulty of the items and the English expression and vocabulary. In doing this, it may be that the items were made too similar.

The original Reading Comprehension test was piloted with sixty students who were either bilinguals or monolinguals. Opinions and suggestions were requested from the students and their teachers. The Reading Comprehension test then was redesigned by taking into consideration the students' and the teachers' opinions. Considered as difficult, the Reading Comprehension test passage was redesigned to contain less unfamiliar vocabulary, and pictures were. This led to the reconstruction of the test with more pictures and repetitive words. Feedback and opinions of students and teachers who preferred more pictures on the reading passage due to its benefit in understanding reading were taken into account. With pictures, the passage became more accessible for the students because the pictures imposed clearer meaning on the content/passage. With repetitive words, the passage became more understandable for the students who appeared to have limited vocabulary. However, by redesigning the test, it seemed that the items were made too similar, which led to the test not having enough difficult items.

It appeared that this redesigned Reading Comprehension passage was not so easy for some students and they could not answer the items as predicted. Some students with a medium-ability could answer the medium items correctly; some had difficulty with the predicted easy items. Similarly, it was predicted that the higher-ability students would be able to answer the easy items and medium difficulty items correctly, but they were unable to do so for all items. This resulted in a less than ideal agreement among the 780 students about the item difficulties along the scale. One reason for the disagreement about the item difficulties was probably guessing. It seemed that some

students still found the items difficult, even though the test had been redesigned prior to the actual test, after the piloting stage, and they guessed the answers to the items that they found difficult.

The presumed guessing that some of the students did is understandable due to their lack of English ability. They had only been learning English as a foreign language for about 30-40 hours in the semester when the test was conducted. Though the curriculum stated that their ability was in line with the Reading Comprehension being tested, much of the time the teaching-learning progress ran much slower than the curriculum suggested. Teachers who were interviewed regarding this matter confessed that they could not run the teaching-learning process, as stated by the curriculum timeline because it would produce even poorer student achievement. The teachers, therefore, needed to pace the students' progress, which meant that some students were behind the curriculum timeline. This could have resulted in students' possessing very limited English when they sat for the Reading Comprehension test for this research. Their lack of ability in English could probably explain the question of why, after redesigning the Reading Comprehension passage test, with more repetitive words and pictures, the students still found the test difficult.

The construct validity of the test was tested by designing the items in ordered patterns of item difficulty which then were compared with their Rasch-measured item difficulties. That is, the Rasch measured item difficulty order given in Table 6.5 was similar to what was predicted when the items were created and designed. The overall fit to the Rasch Measurement Model, as shown by the item-trait interaction chi-square, was $\chi^2=200.4$, $df = 60$ with $p=0.000$, which was not ideal. This meant that the agreement amongst all the students about the item difficulties along the scale was not ideal.

The Cronbach Alpha was 0.53 and low. However, some other measurement aspects were more satisfactory. The thresholds were ordered in line with the scoring categories and the Scoring Category Curves were appropriately ordered with overall measures. The Item Characteristic Curves were satisfactory in most cases, without being ideal, and targeting, while reasonable, would have been improved if there was a wider range in question difficulty. The student measures ranged from -0.8 to +3.9 logits and the item thresholds ranged from -1.2 to +2.6 logits.

Even though the fit to the Rasch Measurement model was not ideal, the Item Characteristic Curves showed no Differential Item Functioning against gender (male versus female) but the Item Characteristic Curves showed Differential Item Functioning against the type of teaching (bilingual versus monolingual teaching). The Differential Item Functioning (DIF) against type of teaching is interpreted as not really being DIF, but as showing that there is a real difference in output by bilingually-taught against monolingual-taught students, with Bilinguals achieving at a higher standard on most items. This is supported by the difference in overall Rasch measures for reading Comprehension by type of teaching. The bilingually-taught mean is 2.094 with $N=394$ and $SD=0.85$ and the monolingually-taught mean is 0.982 with $N=386$ and $SD=1.07$ ($F=257.89$, $df=1,778$, $p=0.00000$). These results showed that the Rasch measure was not ideal and still needed some improvement.

While girls performed significantly higher in English Reading Comprehension than boys overall — in line with the results from many western countries — only two items showed a statistically significant result for girls (item 10 and item 12). Item 10 was a multiple choice and it is hard to see why it should favour girls. Item 12 involved writing and English comprehension which possibly required more thinking and motivation to write to answer the "why" question, and in retrospect, this could be expected to favour girls over boys, not because the item is biased, but because girls in their early secondary school years (12-13 years) were better than boys, in regard to English Reading Comprehension, as shown on items 1, 3, and 6 (see, for example Table 5.9 for Item 1).

In the case of language instruction (bilingual or monolingual instruction), five items showed DIF, with bilingual instruction being superior (items 2, 5, 6, and 9). Items 2, 5, 6, and 9 involved slightly harder reading words and comprehension than the other items and an understanding of their meaning would have been enhanced through bilingual instruction.

By performing DIF against both gender and language instruction, even with a less than ideal fit to the Rasch measurement model, the reason for the DIF could be worked out. Language of instruction would appear to be an important determinant of

performance in English Reading Comprehension and bilingual instruction was related to a superior overall performance compared to monolingual instruction.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for English Reading Comprehension – based on the linear Rasch-created measures.

CHAPTER SIX

DATA ANALYSIS (PART 2)

EXPERIMENTAL RESULTS FOR READING COMPREHENSION

In this chapter, the linear Rasch measures for Reading Comprehension for the bilingually-taught group, known as the experimental group, were compared with similar measures for the monolingually-taught group, known as the control group, in a pretest/posttest, control/experimental group design. In order for the Reading Comprehension measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group to all be on the same scale and comparable, the measures on the pretest have been equated with those on the posttest.

The Mixed Design ANOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant & Tennant, 2007) was conducted to assess the comparison of two different interventions (bilingually-taught program and monolingually-taught program), and of gender on students' scores from the Reading Comprehension Questionnaire for the pretest and posttest. The Two-Way ANCOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant & Tennant, 2007) was employed to test for main and interaction effects with SPSS to compare bilinguals with monolinguals and student gender.

Results on Mixed Design ANOVA (Split-Plot Design ANOVA)

The Mixed Design ANOVA, which is also known as the Split-Plot ANOVA, was conducted to compare the measure of Reading Comprehension Questionnaire of the two language programs (bilingual program and monolingual program) and of gender (female and male). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Reading Comprehension Questionnaire posttest, run following completion of the two months of teaching English as a foreign language. The Reading

Comprehension Questionnaire pretest was conducted prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences.

A mixed between-within-subjects analysis of variance (a.k.a Mixed Design ANOVA or Split-Plot ANOVA) was carried out to assess the difference between two interventions (bilingually-taught program and monolingually-taught program); and the comparison of gender, on participants' Reading Comprehension measures, across two time periods (pretest and posttest).

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity. The result stated that there was no significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.000, $F(1,778) = 0.068$, $p = 0.794$, partial eta squared = 0.000. Wilks' Lambda is a statistical test used in the multivariate analysis of variance to test whether there are differences between the means of identified groups of subjects on a combination of dependent variables. Smaller values of Wilks' lambda indicate greater discriminatory ability of the function (see http://www.blackwellpublishing.com/specialarticles/jcn_9_381.pdf). Partial eta squared (η_p^2) is the default effect size measure in SPSS.

However there was a main effect for time (pretest and posttest), although small, with Wilks' Lambda = 0.985, $F(1,778) = 12.125$, $p = 0.001$, $\eta_p^2 = 0.015$ (small effect) for the Reading Comprehension measure. The main effect comparing the two types of intervention was significant, with $F(1,778) = 369.341$, $p = 0.000$, $\eta_p^2 = 0.322$ (large effect). According to Cohen's effect size guidelines for $\eta_p^2 = 0.322$ were relatively large size effects (Cohen's η_p^2 values: 0.14 = large effect; 0.06 = medium effect; and 0.01 = small effect) (Cohen, 1988; also cited in Burns, 2000). This suggested a real difference in students' ability in English Reading Comprehension with bilingual students achieved better scores than monolingual students (see Figure 6.1 and Table 6.1 and Table 6.2).

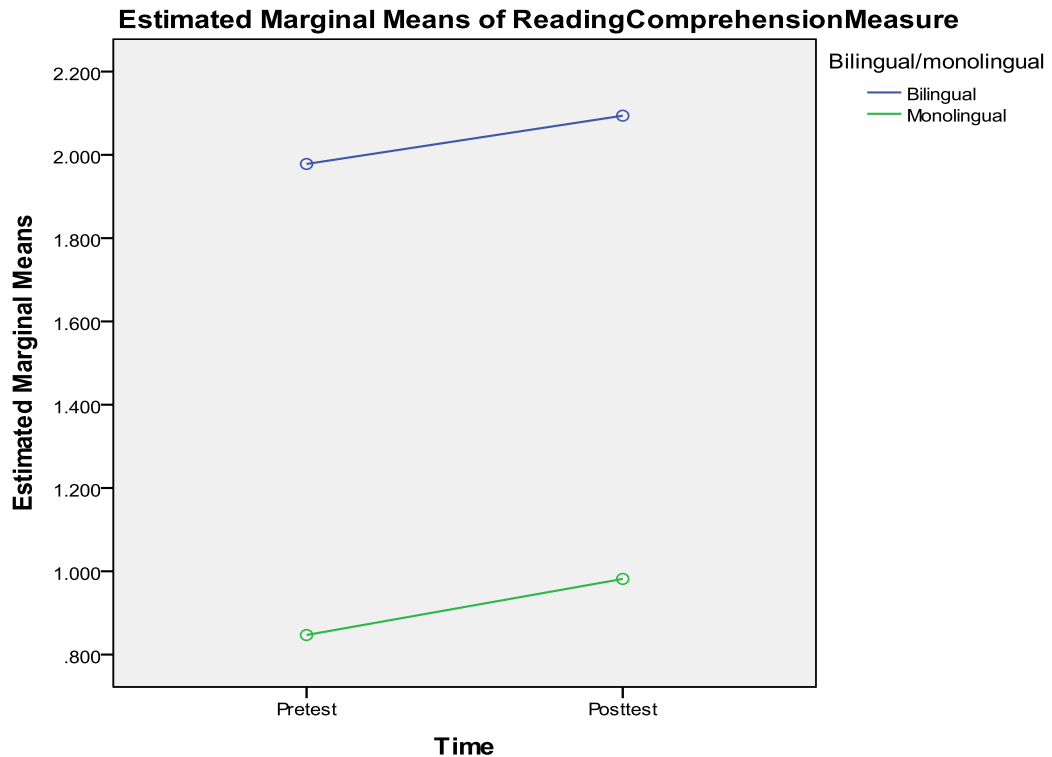


Figure 6.1 Graph of Reading Comprehension by Group

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.999, $F(1,778) = 0.705$, $p = 0.401$, $\eta_p^2 = 0.001$ (a very small effect according to the Cohen's 1988 rules). The main effect for time (pretest and posttest) was not significant either, with Wilks' Lambda = 0.984, $F(1,778) = 12.603$, $p = 0.000$, $\eta_p^2 = 0.016$ (a small effect) for the Reading Comprehension measure. However, the main effect comparing the two types of intervention (control and experimental) was significant, with $F(1,778) = 11.153$, $p < 0.001$, partial eta squared = 0.14 (a large effect size according to Cohen's 1988 regulations). This suggested that there was a real difference in students' ability in English Reading Comprehension, between females and males (see Figure 6.2).

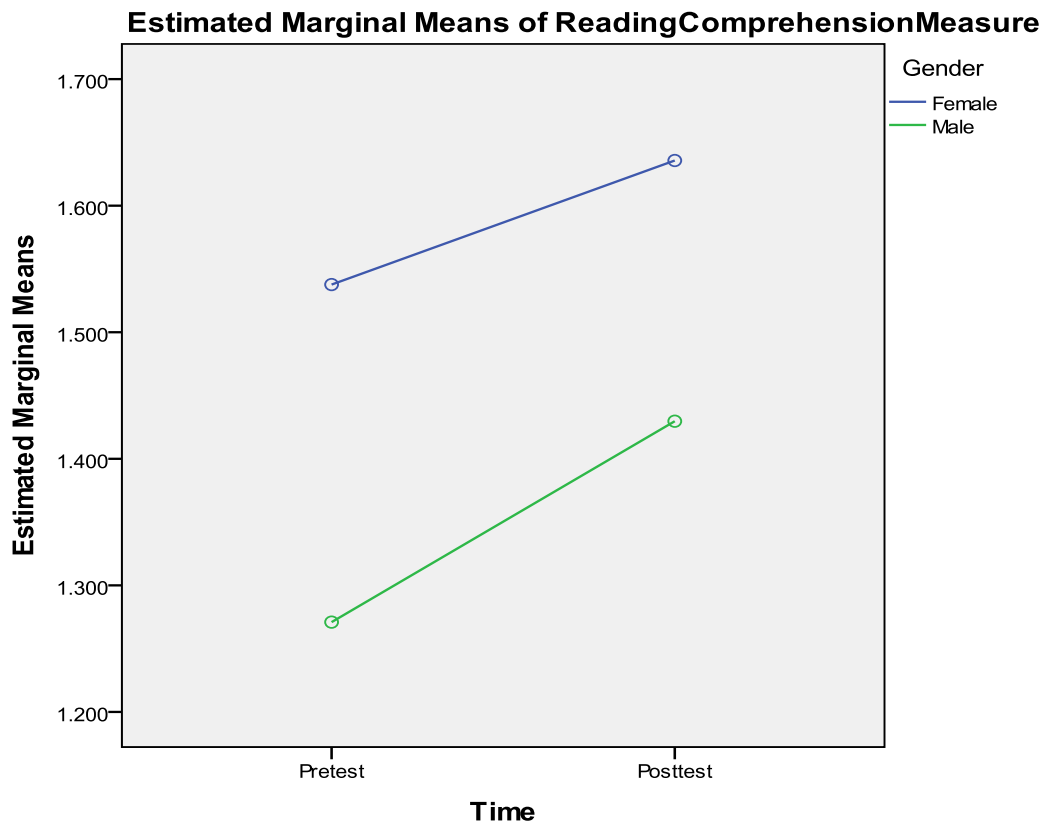


Figure 6.2 Graph of Reading Comprehension by Gender

Table 6.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	1.98	.97	386	.85	.93
Posttest	394	2.09	.85	386	.98	1.07

Table 6.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	1.54	1.03	349	1.27	1.17
Posttest	431	1.64	1.05	349	1.43	1.18

Figure 6.1 and Table 6.1 indicate that bilinguals achieve higher than monolinguals and Figures 6.2, and Table 6.2 indicate that girls have performed moderately better than boys.

Results on Two-Way ANCOVA

The following are the Two-Way ANCOVA results for the Reading Comprehension measures. This test was administered to assess the comparison of the two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught), and gender (male and female). The dependent variable was scores on the posttest, administered following the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language (Time 2). Scores on the pretest administered prior to the start of the programs (Time 1) were used as a covariate to control for individual differences.

Preliminary checks were run to confirm that there were no violations of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for the pretest of Reading Comprehension measure, the result affirmed that there was no significant interaction effect: $F(1, 775) = 0.663, p > .005 (.416)$, with a small effect size ($\eta_p^2 = 0.001$) (see Table 6.3).

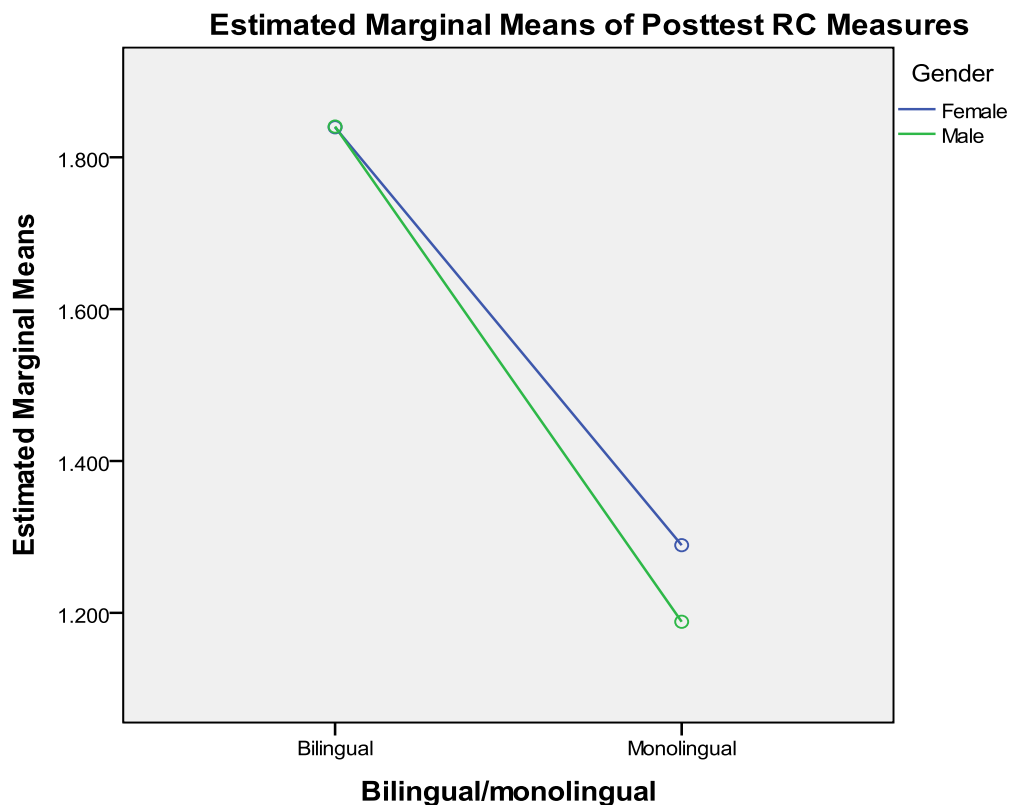
Table 6.3

Interaction Effect and Main Effect of Reading Comprehension (RC) between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest RC Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared (η_p^2)
Corrected Model	389.195 ^a	4	97.299	130.019	.000	.402
Intercept	191.957	1	191.957	256.509	.000	.249
Pretest RC Rasch Measure on Posttest Scale	141.673	1	141.673	189.316	.000	.196
Bilingual monolingual	51.115	1	51.115	68.304	.000	.081
Gender	.473	1	.473	.633	.427	.001
Bilingual monolingual *						
Gender	.496	1	.496	.663	.416	.001
Error	579.965	775	.748			
Total	2827.510	780				
Corrected Total	969.160	779				

a. R Squared = .402 (Adjusted R Squared = .398)

Note: the '*' indicates interaction

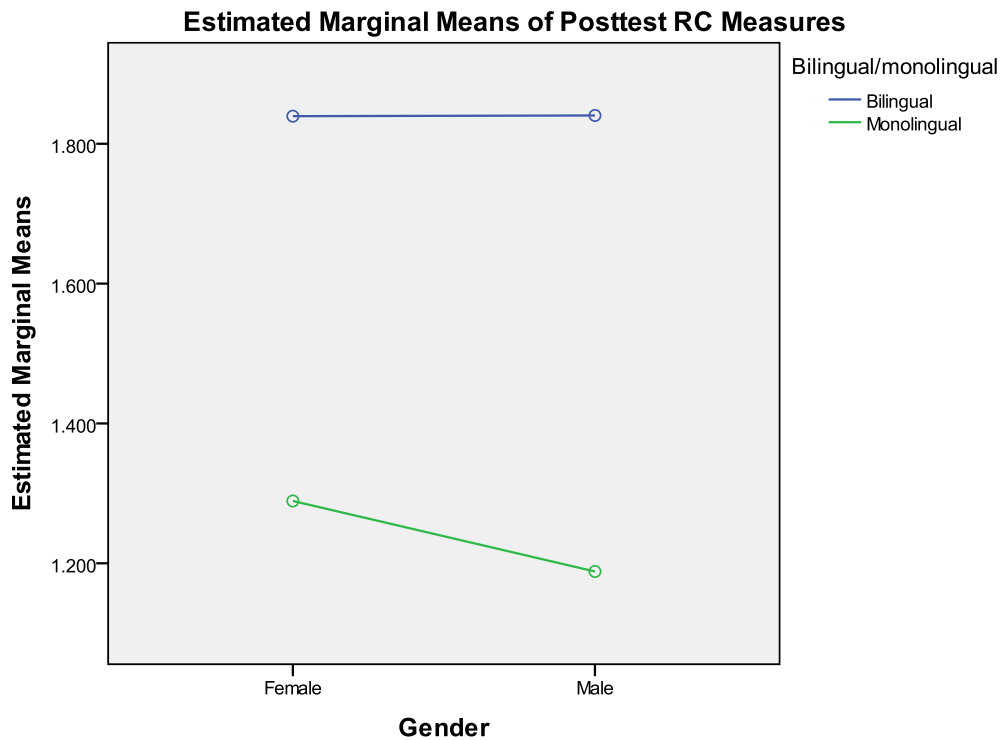


Covariates appearing in the model are evaluated at the following values: Pretest RC Rasch Measure on Posttest Scale = 1.41833

Figure 6.3 Graph of Reading Comprehension by Group (Bilingual/monolingual) with Pretest as Covariate

The main effect for the language program was statistically significant with $F(1,775) = 68.304, p = 0.000$, with $\eta_p^2 = 0.81$ (a large effect size according Cohen's 1988 regulations); while that for gender was not, with $F(1,775) = 0.633, p > 0.005 (0.427)$ with $\eta_p^2 = 0.001$ (a small effect size) (see Figures 6.3 and 6.4).

These results reveal that female students' scores on the Reading Comprehension was equally better than those of male students; and female and male students in the bilingual program responded equally well to the female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest RC Rasch Measure on Posttest Scale = 1.41833

Figure 6.4 Graph of Reading Comprehension by Gender with Pretest as Covariate

Summary

Bilingual (N = 394) and monolingual (N = 386) students were given an English Reading Comprehension pretest and posttest within a two month period (equivalent to

16 lessons), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA using the SPSS computer program. There are two main sets of findings: (1) with regards to the Mixed Design ANOVA, and (2) with regards to the Two-Way ANCOVA.

The main findings for English Reading Comprehension, with regards to the Mixed Design ANOVA, are:

1. There was no significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.000, $F(1,778) = 0.068$, $p = 0.794$, $\eta_p^2 = 0.000$ (small effect).
2. There was a small main effect for time (pretest and posttest), with Wilks' Lambda = 0.985, $F(1,778) = 12.125$, $p = 0.001$, $\eta_p^2 = 0.015$ (small effect) for the Reading Comprehension measure.
3. The main effect comparing the two types of intervention was significant, with $F(1,778) = 369.341$, $p = 0.000$, $\eta_p^2 = 0.322$ (large effect).
4. For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.999, $F(1,778) = 0.705$, $p = 0.401$, $\eta_p^2 = 0.001$ (small effect).
5. The main effect for time (pretest and posttest) was not significant, with Wilks' Lambda = 0.984, $F(1,778) = 12.603$, $p = 0.000$, $\eta_p^2 = 0.016$ (small effect)
6. The main effect comparing the two types of intervention was significant, with $F(1,778) = 11.153$, $p < 0.001$, $\eta_p^2 = 0.14$ (large effect).

The main findings for English Reading Comprehension, with regards to the Two-Way ANCOVA, are:

1. There was no significant interaction effect: $F(1, 775) = 0.663$, $p > .005$ (.416), with $\eta_p^2 = 0.001$ (small effect).
2. The main effect for the language program was statistically significant with $F(1,775) = 68.304$, $p = 0,000$, with $\eta_p^2 = 0.81$ (a large effect size according Cohen's 1988 regulations).

3. The main effect for the gender was not significant with $F(1,775) = 0.633$, $p > 0,005$ (0.427) with $\eta_p^2 = 0.001$ (a small effect).

The findings indicated that female and male bilingual students achieved better scores of Reading Comprehension than female and male monolingual students did. Female students' scores on the Reading Comprehension were better than those of male students.

The next chapter explains the Rasch analysis with the RUMM2030 computer program for the English Writing Test.

CHAPTER SEVEN

DATA ANALYSIS (PART 3)

RASCH MEASUREMENT OF ENGLISH WRITING

This chapter reports the findings of the Rasch analysis of the Writing posttest data.

The Extended Logistic Model of Rasch (Andrich, 1988a) was used to analyse the Writing Test data (I= 12 ‘items’) in order to create a linear measure of students’ English Writing at the posttest stage of the experiment. Each student’s writing was initially marked in four response categories: ‘poor’ (scored 1), ‘adequate’ (scored 2), ‘good’ (scored 3), and ‘excellent’ (scored 4). These three aspects were; (1) paragraph organization (easiest); (2) text conventions (harder); and (3) text quality (hardest). These aspects are called ‘items’ (see Tables 7.1 and 7.2) in order to be compatible with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010). For the rest, the ‘item’ is called item (without brackets).

Initial Analysis (Posttest)

The analysis involved running the data through the RUMM2030 program and checking for non-fitting items and items with poor discrimination, which were deleted, and then re-running the program. After four analyses, six of the original 18 items were rescored (Items 1, 2, 3, 10, 11, and 12); and another six (items 1, 4, 10, 13, 15, and 17) were deleted, leaving 12 items in the final scale: Items 2, 3, 5, 6, 7, 8, 9, 11, 12, 14, 16, 18. The six deleted items did not have a good fit to the measurement model, for two reasons. First, there were overlapping thresholds and poor Scoring Category Curves. Second, the item residuals were too high. Item residuals are differences between the predicted scores, based on the Rasch parameters, and the actual scores on the items. The RUMM program does not tell the researcher why an item does not fit the Rasch Measurement Model but one can infer that many students with the same measures could not agree on the difficulty of the items all along the scale.

A more detailed reason for why some items did not fit the Rasch model was difficult to determine. However, it was assumed that a substantial cause of the misfit was the students' lack of ability in the writing test. The students were in their first semester of the first year in which English was being taught for the first time. They were still weak in English in general and in English writing in particular. The misfit may have been primarily due to their low command of English writing related to the different classroom cultures (Indonesian's way of writing is different from that of English's way) in the two groups (bilingually-taught and monolingually-taught), in a test situation. Thus, it may be that the two groups of students (bilingually-taught and monolingually-taught) did not agree on the difficulty of the items.

The English writing test consisted of two topics. The first topic was 'My *Idul Fitri* Holiday' and the second one was 'My Family'. The difficulties of the items on the two topics was predicted before data collection, based on logic and experience and the 18 items were placed on an ordered scale (see Tables 7.1 and 7.2). Paragraph Organisation were ordered from easy to hard for Topic Sentence, Concluding Sentence and Supporting Sentences); Text Convention was ordered from easy to hard for spelling, punctuation/capitalisation, and grammar); and Text Quality was ordered from easy to hard for readability, style, and text enjoyability (see Appendix D for the scoring rubric).

The three levels of difficulty were ordered from paragraph organisation (considered the easiest), to text conventions (considered to be hard), and text quality (expected to be harder). For Acehnese students learning English writing for the first time in their first semester of Middle Schools, it was considered that their previous knowledge and experience in Indonesian writing has helped them in understanding and ability in English writing. It was expected that the students would have acquired some paragraph organisation from their previous teaching and learning in Indonesian and, therefore, paragraph organisation was expected to be the easiest, because paragraph structure is similar in Bahasa Indonesia and English.

As students who have just started their English lessons, it was expected that they might find text conventions more difficult than paragraph organisation. The students would have had a small exposure to English and they not only lack depth in

vocabulary, but also in spelling, punctuation/capitalisation, and grammar. Thus, text conventions were considered to be harder than the paragraph organisation.

It was expected that students would find text quality the most difficult in their English writing and that they would need substantial English writing experience to do this well. Some students might be able to acquire paragraph organisation in which they are able to produce paragraphs with a topic sentence, some supporting sentences, and finish with a concluding paragraph. They might also be able to make their writing understandable by using the correct vocabulary, spelling, punctuation, capitalisation, and grammar. But, in the early stages, they generally find it difficult to ensure that the sentences are linked together so that the information ‘flows’ nicely for a reader, and that their writing has some style, so that it can be enjoyed by the reader. Therefore, the text quality was considered the most difficult of all.

Each topic (either My *Idul Fitri* Holiday or My Family), has an order of difficulty of ‘easy’, ‘hard’ and, ‘harder’. For ‘Paragraph Organisation’, the ‘main topic’ was considered easy because it involved simple knowledge. The ‘Concluding Sentence’ was considered ‘harder’ because it involved a little more than simple knowledge. The ‘Supporting Sentences’ were considered ‘harder’ as they involved a higher order of thinking which students needed more practice in and knowledge of in order to write around seven or eight correct sentences.

For ‘text conventions’, ‘spelling’ was considered ‘easy’, because it involved common knowledge of the spelling of a noun, verb, adjective or adverb in English. In Indonesia, the spelling of word classes is among the first aspects taught. ‘Punctuation/capitalisation’ was considered ‘hard’ because it involved ability to recognise punctuation/capitalisation which might not be easy for students to do. ‘Grammar’ was considered the hardest compared with ‘spelling’ and ‘punctuation/capitalisation’, because it involved a higher order of ability and knowledge.

For ‘text quality’, ‘readability’ (i.e. clear and neat handwriting) was considered ‘easy’ because it involved a simple ability that students had already achieved. ‘Style’ (i.e. sentence fluency, for example varied length, good flow and rhythm, and varied structure) was considered ‘hard’ because it involved more ability and skills. ‘Text

enjoyability’ was considered the hardest of all under ‘text quality’ because it involved more advanced ability in order to make readers enjoy the writing.

Marks for the writing were allocated from the highest (4) to the lowest (0), based on the English writing scoring rubric (see Table 3.2). Each piece of writing was marked on the 9 items (for each topic) based on the English writing scoring rubric (see Table 3.2). Table 7.1 is an example of the marking used for Writing.

Table 7.1
Writing Marking Example for Topic 1 ‘My Idul Fitri Holiday’

Student									
ID	Item 1 (Title: My Idul Fitri Holiday)								
	Topic. Sent.	Conc. Sents.	Supp. Sents.	Spelling	Punc./ Cap.	Grammar	Read.	Style.	Text Enj.
001	3	1	4	2	2	2	3	3	2
002	1	1	4	3	3	2	3	3	2
003	3	2	4	3	3	2	3	3	3
004	1	2	4	3	3	2	3	3	3
005	3	1	4	2	2	2	3	2	2
006	3	2	4	3	3	2	2	3	2
007	3	2	4	3	3	2	3	3	2
008	3	3	4	3	2	3	3	2	3
009	3	1	4	3	3	2	3	3	2
010	3	1	4	3	3	2	3	2	2

The original behaviour of 18 items (item difficulties for both writing topics) for the difficulties prior to the Rasch analysis and after the Rasch analysis differs. This is due to the fact that the 18 item difficulties after the analysis were fitted to the Rasch model. Table 7.2 shows the original of 18 item difficulties prior to the analysis (with ‘Topic 1 My Idul Fitri Holiday, topic sentence the easiest, and Topic 2 My Family, text enjoyability, the most difficult).

Table 7.2
Initial Marking of English Writing Essay as ‘Items’ (Scored in terms of 18 Items for Compatibility with RUMM2030)

No.	Original Number	
1	1	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Topic Sentence
2	2	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Concluding Sentence
3	3	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Supporting Sentences
4	4	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Spelling
5	5	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Punctuation and Capitalisation
6	6	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Grammar
7	7	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Readability
8	8	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Style
9	9	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Text Enjoyability
10	10	Topic 2 ‘My Family’ Topic Sentence
11	11	Topic 2 ‘My Family’ Concluding Sentence
12	12	Topic 2 ‘My Family’ Supporting Sentences
13	13	Topic 2 ‘My Family’ Spelling
14	14	Topic 2 ‘My Family’ Punctuation and Capitalisation
15	15	Topic 2 ‘My Family’ Grammar
16	16	Topic 2 ‘My Family’ Readability
17	17	Topic 2 ‘My Family’ Style
18	18	Topic 2 ‘My Family’ Text Enjoyability

Note: The item difficulties for the Writing test have been ordered from Topic 1 ‘My *Idul Fitri* Holiday’ to Topic 2 ‘My Family’, and scoring is set out in terms of ‘items’ for compatibility with the RUMM2030 computer program.

Unlike Table 7.2 that shows the original of 18 item difficulties, before being analysed by RUMM230, Table 7.3 shows the 12 marked items whose data fit the Rasch measurement model, with Topic 1 My *Idul Fitri* Holiday concluding sentence as the easiest marked item and Topic 2 My Family text enjoyability as the most difficult marked item.

Table 7.3

Final Marking of English Writing Essay as Items (Scored in terms of 12 Items for Compatibility with RUMM2030)

No.	Original Number	
1	2 (easiest)	Topic 1 'My <i>Idul Fitri</i> Holiday' Concluding Sentence
2	3	Topic 1 'My <i>Idul Fitri</i> Holiday' Supporting Sentences
3	5	Topic 1 'My <i>Idul Fitri</i> Holiday' Punctuation and Capitalisation
4	6	Topic 1 'My <i>Idul Fitri</i> Holiday' Grammar
5	7	Topic 1 'My <i>Idul Fitri</i> Holiday' Readability
6	8	Topic 1 'My <i>Idul Fitri</i> Holiday' Style
7	9	Topic 1 'My <i>Idul Fitri</i> Holiday' Text Enjoyability
8	11	Topic 2 'My Family' Concluding Sentence
9	12	Topic 2 'My Family' Supporting Sentences
10	14	Topic 2 'My Family' Punctuation and Capitalisation
11	16	Topic 2 'My Family' Readability
12	18 (hardest)	Topic 2 'My Family' Text Enjoyability

Note. The misfitting 'items' are 1, 4, 10, 13, 15, and 17 and they were deleted.

The following the output from the RUMM2030 program supports the creation of a linear scale of Writing English with 12 'items'.

Output from the Final Rasch Analysis (Posttest)

Standardised Fit Residuals

In an ideal measurement scale, the standardised fit residuals should yield an approximately normal distribution with mean value = 0 and standard deviation value = 1, with minimal differences between the actual and expected test score values. The fit residual data for students of the Writing test had a mean near zero ($M = -0.35$) and standard deviation near 1 ($SD = 0.7$) and for items the values were $M = -1.89$ and $SD = 2.1$. This means that the student measure-item response pattern was not ideal and that some revision of the marking scheme is probably needed in any future use of the scoring rubric. Table 7.4 shows item and person overall fit statistics for the Writing test.

Table 7.4
Overall Fit Statistics for the Writing Test (N=780, I=12)

ITEM-PERSON INTERACTION				
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-1.888	-5.243	-0.346
Standard Deviation	3.404	2.094	3.644	0.701

Notes.

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The standardized fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model. They are satisfactory for the students (persons) but not ideal for the 'items'.

Dimensionality and Item-Trait Interaction

The item-trait interaction chi-square was 487.25, $df=108$, and $p=0.00$ (see Table 7.5). This shows that there was a significant interaction between the students and the scores on the items in the Writing test, indicating that there was not an ideal agreement amongst the students about the item difficulties all along the scale. This means that an accurate prediction of each student's scoring on each item could not be achieved using a single parameter for each student (that is the person measure), and a single parameter for each item (that is the item difficulty), as required for a good fit to the Rasch measurement model. While 'how accurate' one needs is a matter of conjecture, how one could obtain a better fit to the measurement model for English as a second language for Indonesian students should be researched in the future.

A factor analysis of the components of the residuals supports the view that the measure is not unidimensional and that there are at least two factors present. The first eigenvalue is 2.9, and it is above the chance level (see <http://www.rasch.org/rmt191h.htm>). Eigenvalues provide information about how much discriminating ability a function possesses.

Table 7.5
Item-Trait Interaction for the Writing Scale

Total Item Chi-Square	487.25
Separation Index (Reliability)	0.94
Total Degree of Freedom	108.00
Total Chi-Square Probability	0.00
Cronbach Alpha (Reliability)	0.95

Notes

1. The Index of Person Separation (Andrich & Van-Schoubroeck, 1989) is interpreted like a Cronbach Alpha (Cronbach, 1951) and is high and good. Cronbach Alpha is based on the raw scores and the Person Separation Index is based on the Rasch parameters.
2. The item-trait interaction test is a chi-square. The results indicate that the collective agreement amongst the students about the item difficulties along the scale was not ideal.
3. All numbers are given to two decimal points because the errors are only up to two decimal points.

Reliability

The Person Separation Index (PSI) is the standard reliability measure based on the Rasch parameters. It is 0.94 (see Table 7.5 above) and very satisfactory, indicating that there is a reliable separation of student measures in relation to the errors all along the scale. Cronbach Alpha (Cronbach, 1951) the standard reliability measure based on the raw scores, is 0.95 and very satisfactory.

Item Statistics

The RUMM 2030 program calculates the item statistics based on the Rasch parameters. Table 7.6 presents the item difficulties, the standard errors, and the item residuals. While some residuals are satisfactory, others are not ideal, showing that not all items fit the measurement model as well as one would like. It is generally expected that residuals would fall within the range -2.5 to + 2.5 but, as shown in Table 7.6, some residuals were outside this range. Deleting these items and re-analysing the data did not improve the fit to the measurement model.

Table 7.6
Locations, Standard Errors, Residuals and Chi-Squares for Writing ‘Items’

Item	Location	SE	Residual
12	-5.499	0.089	1.361
3	-5.291	0.091	-0.353
16	-1.714	0.101	-1.733
7	-1.662	0.102	-2.508
14	-0.864	0.095	-1.939
5	-0.699	0.098	-4.406
8	0.216	0.097	-3.980
18	0.417	0.101	-2.808
9	2.384	0.103	-3.704
6	2.940	0.115	-4.206
11	4.824	0.172	1.078
2	4.947	0.187	0.541

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.

Scoring Category Curve

The RUMM 2030 program produces Scoring Category Curves that examine scoring reliability and show the relationship between the probabilities of scoring in each category as the measure increases. Each item was originally scored in four categories: poor (scored 0), adequate (scored 1), good (scored 2) and excellent (scored 4). For the Writing test, some items were rescored and were marked with only three scoring categories, while the rest were marked with four scoring categories. This was to ensure a high reliability and consistency of marking. The Scoring Category Curves should show a consistent relationship between the probability of scoring in a particular category and the measure, indicating that the scoring was done consistently and logically.

A Scoring Category Curve for item 11 (with three scoring categories) is shown in Figure 7.1. There is a high probability of scoring category 0 and a low probability of scoring category 1, at the lowest student measures. The probability of scoring category 0 decreases and the probability of scoring category 1 increases, as the student measures increase. The probability of scoring category 1 decreases and the probability of scoring category 2 increases, at the highest student measures. The Scoring Category Curves for

the other items were checked and they showed a logical and consistent scoring as well, but they are not shown here.

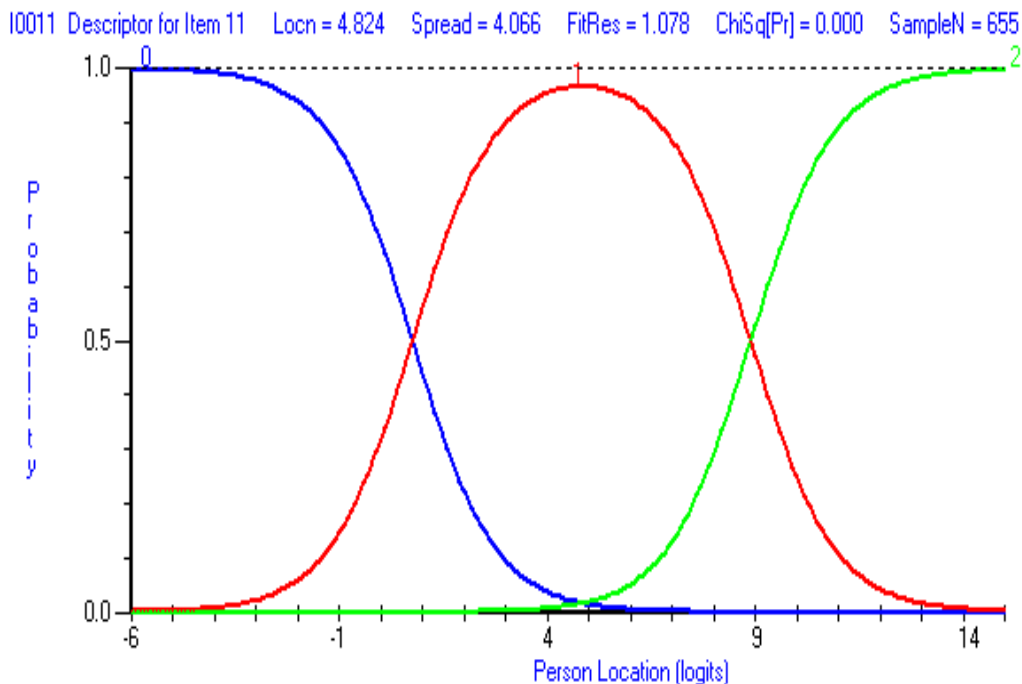


Figure 7.1 Curve for Item Category for Item 11

Note: The blue curve is for a score of 0, the red curve for a score of 1 and the green curve for a score of 2.

Item Characteristic Curves

The Item Characteristic Curve for item 3 (good fitting item) of the Writing test is shown on Figure 7.2. The ogive curve indicates the expected score for the English Writing groups, ranging from the lowest to the highest ability groups. The dots indicate the observed mean measures for the ten student ability groups. When the observed scores closely follow the curve of expected values, the groups are performing as expected on the item. It shows that the item is discriminating well and that as students with higher measures answer the item, they have higher expected values.

The Item Characteristic Curve for item 9 (a not ideal fitting item) of the Writing test is shown on Figure 7.3. The Item Characteristic Curves for the other items were checked. While some were not ideal, they were found to be satisfactory.

10003 Descriptor for Item 3 Locn = -5.291 Spread = 1.814 FitRes = -0.353 ChiSq[Pr] = 0.139 SampleN = 655

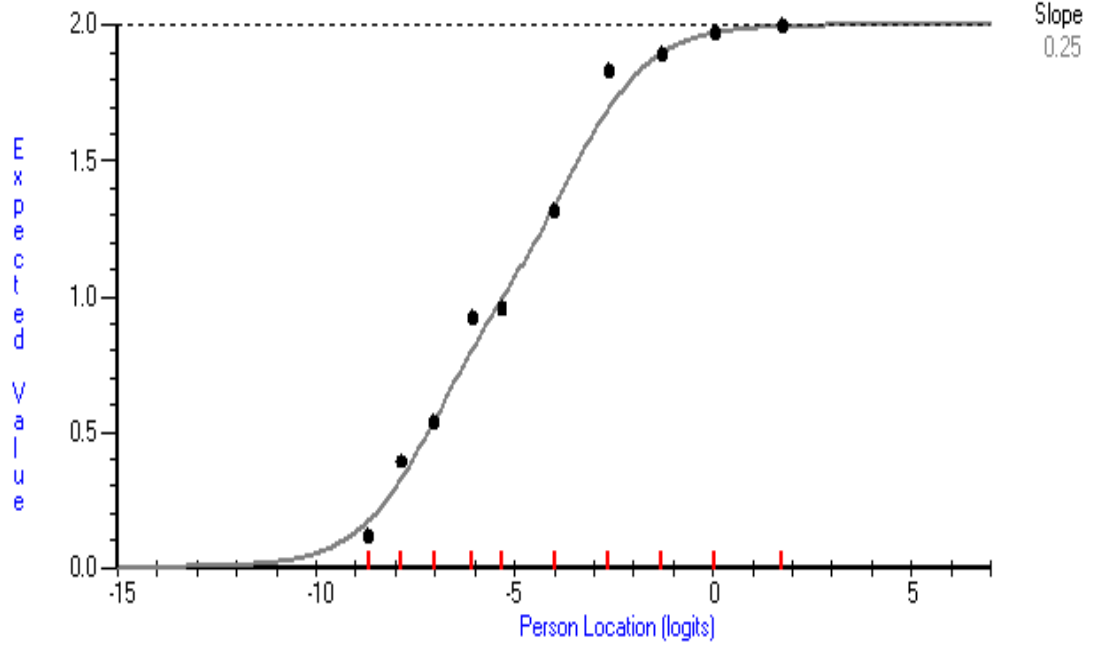


Figure 7.2 Item Characteristic Curve for Item 3

10009 Descriptor for Item 9 Locn = 2.384 Spread = 3.876 FitRes = -3.704 ChiSq[Pr] = 0.000 SampleN = 655

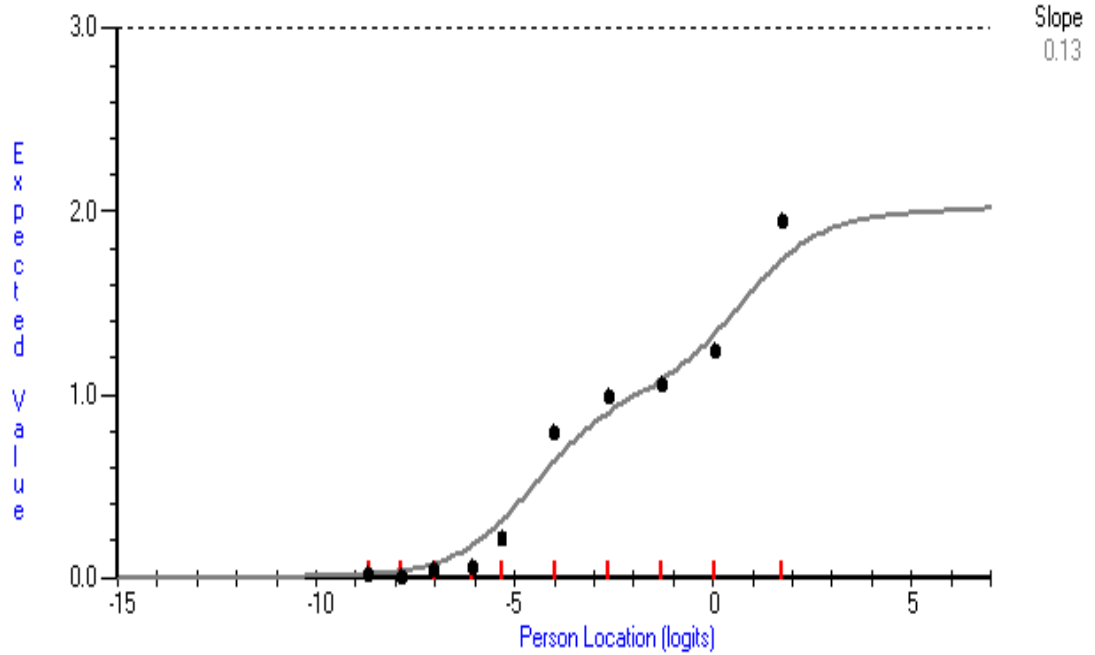


Figure 7.3 Item Characteristic Curve for Item 9 (Not Ideal but Satisfactory)

Person-Item Threshold Distribution (Targeting)

A Person-Item graph (see Figure 7.4) shows the student measures (on the upper side of the scale) and the item difficulties (on the lower side of the scale) from easy (on the left) to hard (on the right). The graph shows that the student measures ranged from -10.5 logits (lowest) to the +6.5 logits (highest), and the item difficulties ranged from -8.5 logits (easiest) to the +10.5 logits (hardest). Measures are considered well-targeted when the range of item thresholds and the range of student measures are about the same. This means that the items were not too hard and not too easy for the students. For this measure, the range of the student measures covered a good range of item difficulties, and so the targeting was very satisfactory. However, some easier items could be added to the scale to improve targeting and some of the most difficult 'items' could be deleted for these students in any future use of the scale.

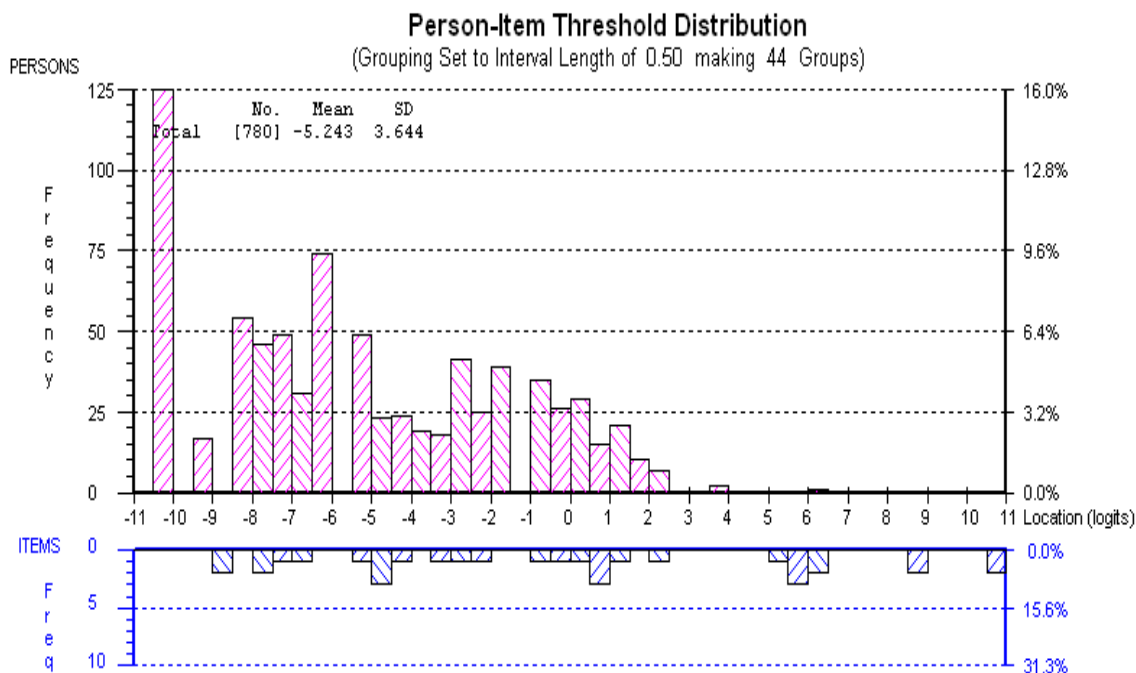


Figure 7.4 Targeting Graph: Student Measures in English Writing (upper-side) and Item Difficulties (lower-side) Calibrated on the Same Scale

Figure 7.4 shows the student measures and the item threshold locations on the same linear scale. Item thresholds are grouped by their locations (difficulties), for example, the locations of 18.2, 9.2, 11.1, and 2.1 are from 0.50 logits to 2.00 logits. The

columns (pink) on the upper-side refer to student measures, which are placed on the same linear scale as the item threshold difficulties, which are placed on the lower-side.

Item thresholds are points between the scoring categories where there is an equal chance of having scores in adjacent categories. In good, consistent measures, the thresholds for the items should be ordered in line with the ordering of the scoring categories, as was the case with this measure (see Figure 7.5).

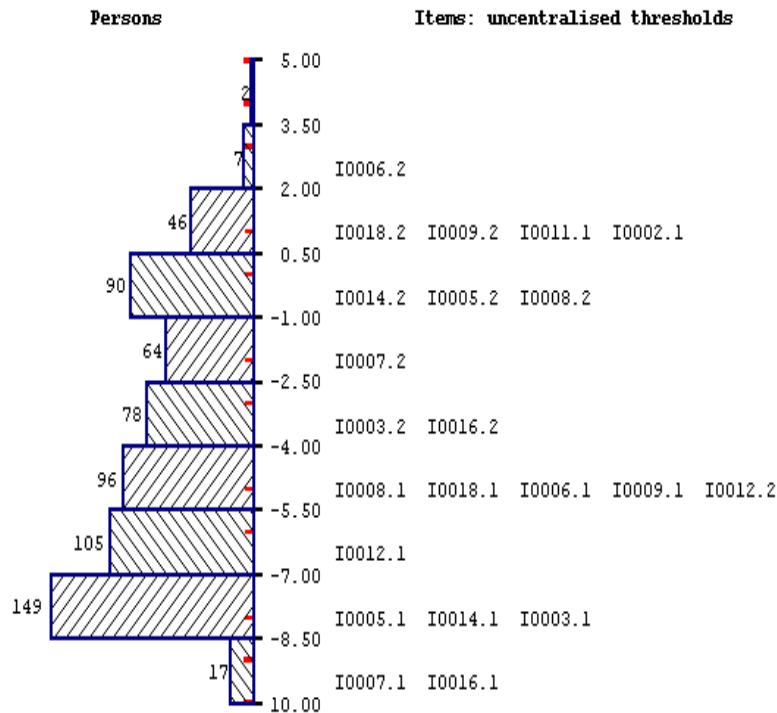


Figure 7.5 Writing Map

Note: I0018.2 means threshold 2 for item 18; I0007.1 means threshold 1 for item 7, and so on.

DIF by Gender

Each of the 12 items of the Writing test showed no statistically significant Differential Item Functioning (DIF) by gender. Item 12 and Item 3 are shown as examples (see Figures 7.6 and 7.7).

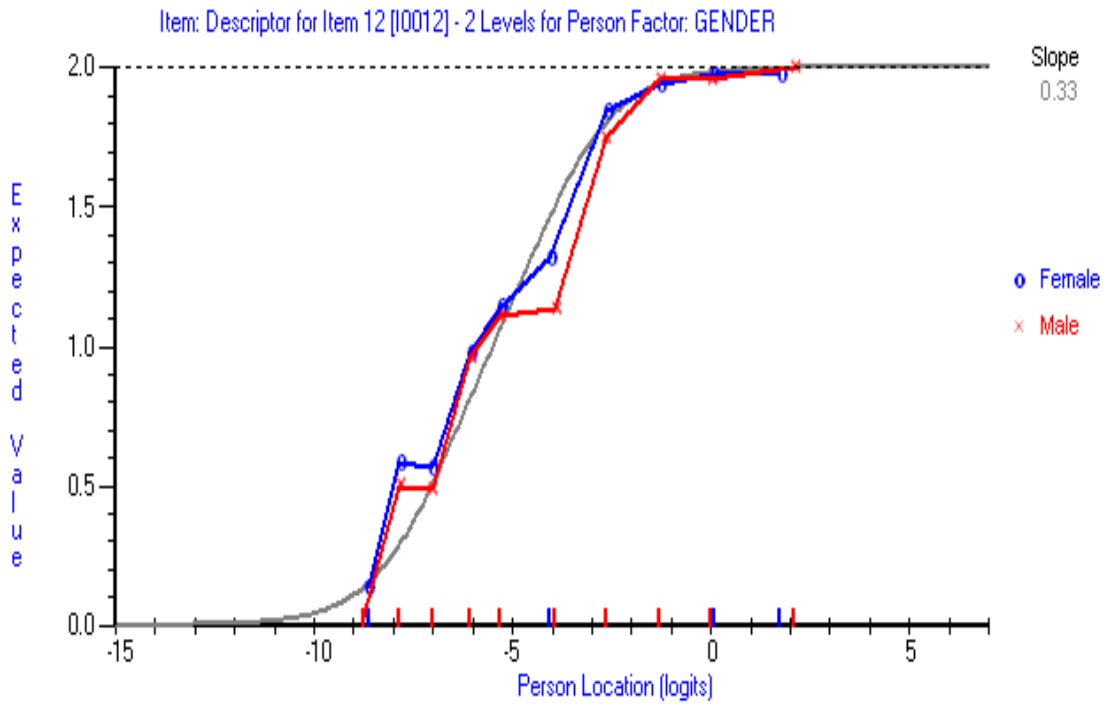


Figure 7.6 Item Characteristic Curve by Gender for Writing Item 12

Note: No statistically significant main effect by gender (no gender bias), $F=1.20$, $df=19,1$, $p=0.27$.

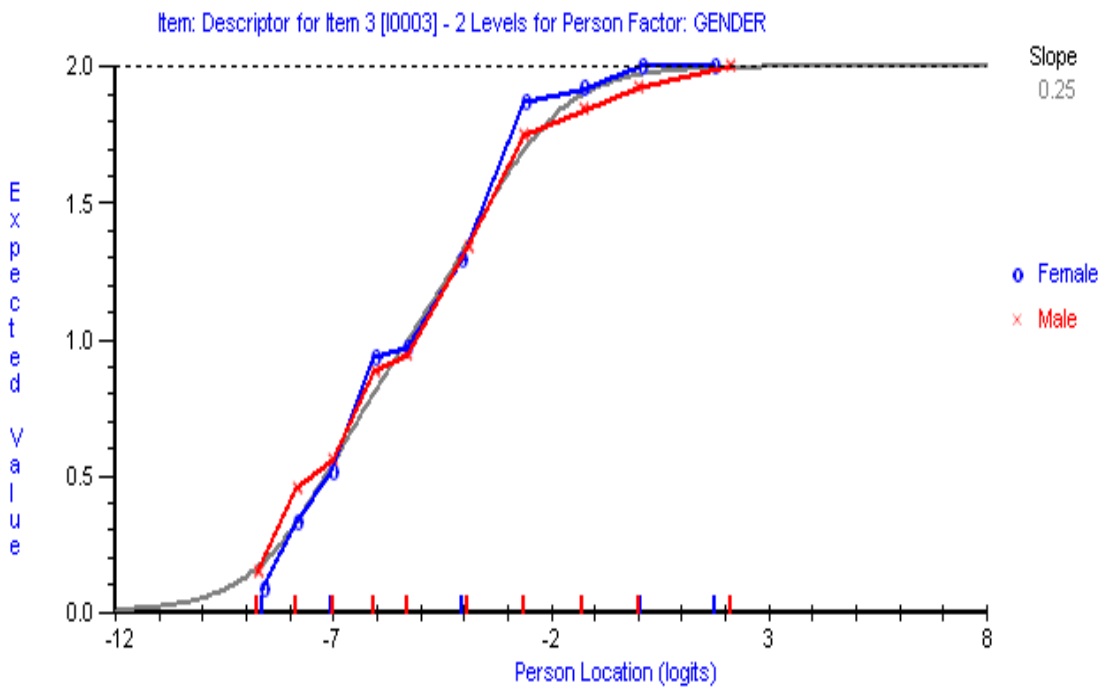


Figure 7.7 Item Characteristic Curve by Gender for Writing Item 3

Note: No statistically significant main effect by gender, no gender bias, $F=1.75$, $df=1,19$, $p=0.18$.

All the Item Characteristic Curves for the other items were checked and showed no statistically significant bias for gender.

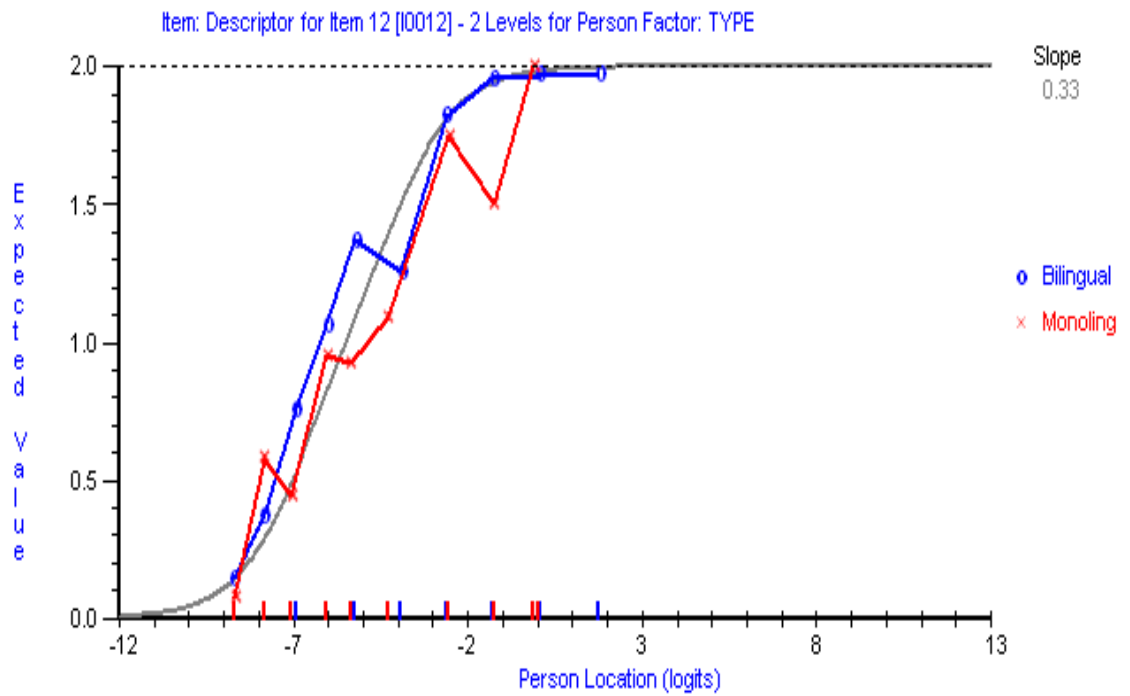


Figure 7.8 Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 12.

Note: No statistically significant main effect by Type of Instruction (no type of language instruction bias), $F=0.004$, $df=1,18$, $p=0.95$

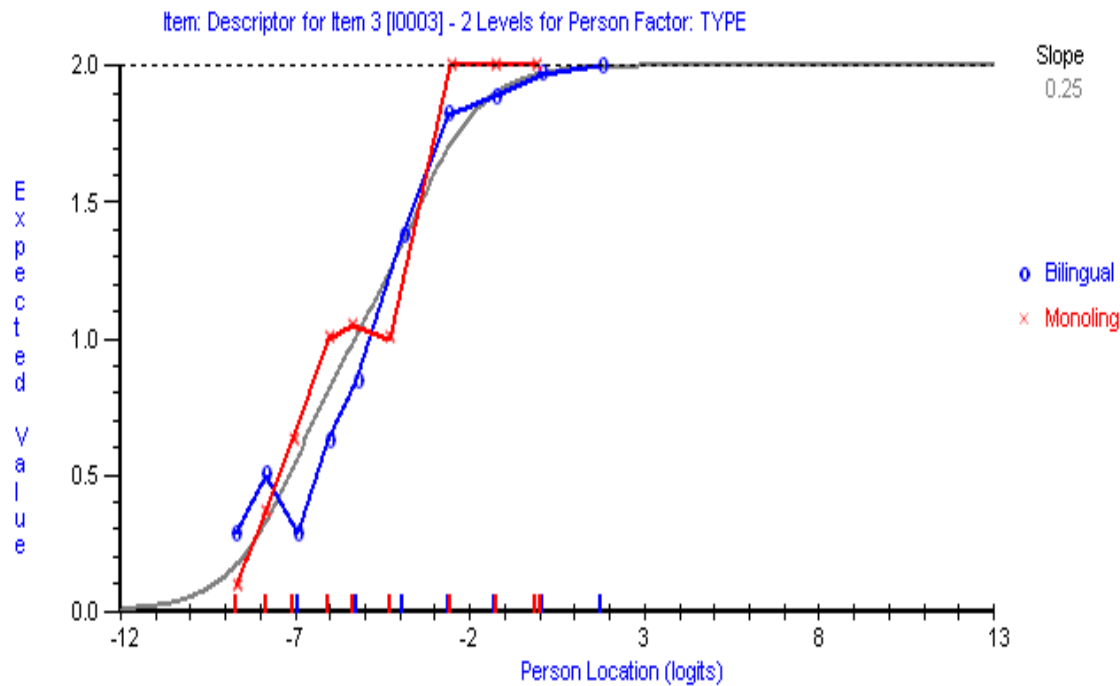


Figure 7.9 Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 3

Note: No statistically significant main effect by Type of Instruction (no type of language instruction bias), $F=1.67$, $df=1,18$, $p=0.20$

DIF by Type of Language Instruction

Each of the 12 items of the Writing test showed no statistically significant Differential Item Functioning (DIF) by Type of Language Instruction (bilingual versus monolingual). Item 12 and Item 3 were taken as examples (see Figures 7.8 and 7.9). This shows that there was no inherent bias in the items with regard to either of the language instruction types.

Targeting by Gender and Type of Language Instruction

Girls were statistically significantly better in Writing than boys ($F= 59.39$, $df=1,778$, $p=0.00000$, see Figure 7.10) and bilinguals did statistically significantly better in Writing than students in monolingual schools ($F= 706.45$, $df=1,778$, $p=0.00000$, see Figure 7.11).

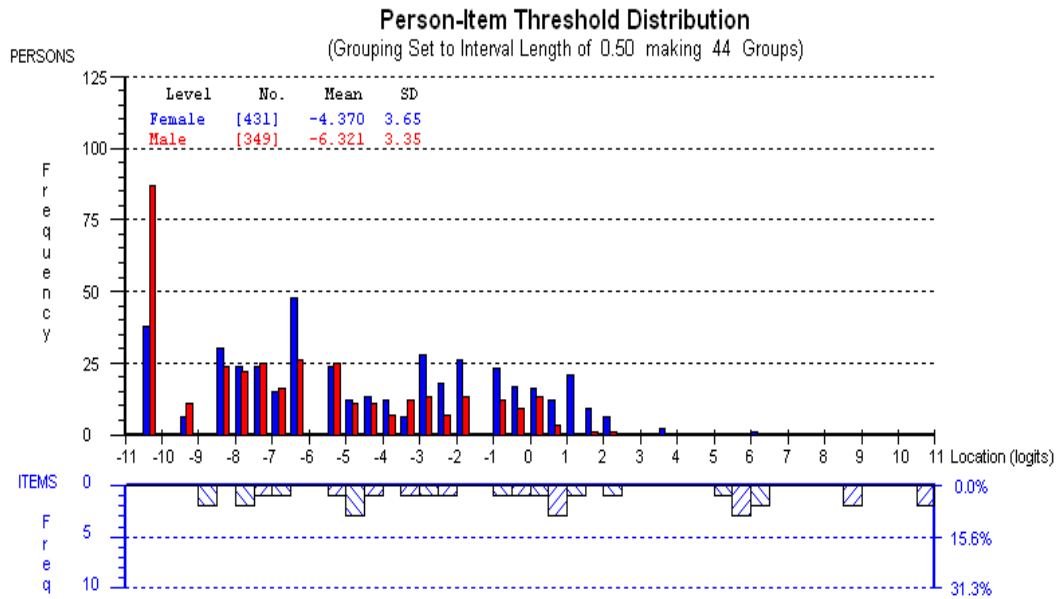


Figure 7.10 Targeting of Writing by Gender

Notes:

1. The person measures are on the upper-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the lower-side of the graph from easy (LHS) to hard (RHS). $F=59.39$, $df=1,778$, $p=0.00000$, meaning that the difference between females and males is statistically significant. Females performed better than males.

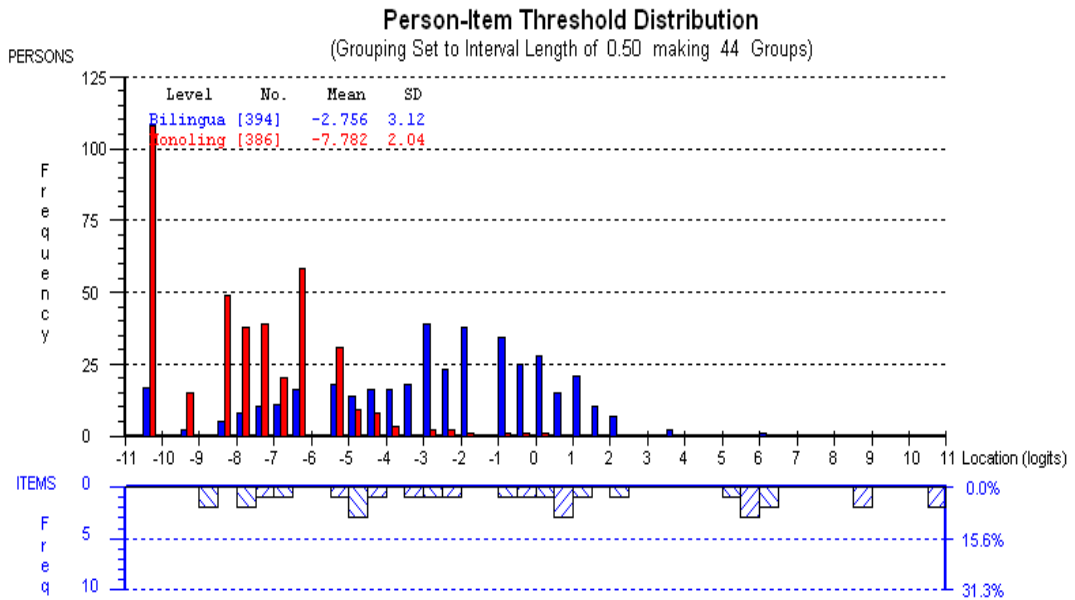


Figure 7.11 Targeting of Writing by Type of Language Instruction

Note: $F=706.45$, $df=1,778$, $p=0.00000$, meaning that the difference between bilinguals and monolinguals is statistically significant. bilinguals performed better than monolinguals.

Scale of Item Difficulties

The corresponding item difficulties for both topics were very similar, showing support for the scoring method and the item difficulty prediction. Table 7.7 shows the items for the Writing Test which have been ordered by difficulty from the easiest to the most difficult on the linear Rasch-created scale. The ordering of the items by difficulty is consistent with the initially predicted conceptualized order, giving support for the construct validity of the scale.

Table 7.7
Ordered Difficulty of Items on the Linear Scale

No	Item Location	Item Statements on the Writing
Easy		
12	-5.50	Topic 2 My Family Supporting Sentences
3	-5.29	Topic 1 My Idul Fitri Holiday Supporting Sentences
16	-1.71	Topic 2 My Family Readability
7	-1.66	Topic 1 My Idul Fitri Holiday Readability
14	-0.86	Topic 2 My Family Punctuation/Capitalisation
5	-0.70	Topic 1 My Idul Fitri Holiday Punctuation/Capitalisation
8	0.22	Topic 1 My Idul Fitri Holiday Style
18	0.42	Topic 2 My Family Text Enjoyability
9	2.38	Topic 1 My Idul Fitri Holiday Text Enjoyability
6	2.94	Topic 1 My Idul Fitri Holiday Grammar
11	4.82	Topic 2 My Family Concluding Sentence
2	4.95	Topic 1 My Idul Fitri Holiday Concluding Sentence
Hard		

Note: Item difficulties (locations) are measured in logits, the log odds of answering successfully.

The items were ordered from easy to hard on a linear scale (see Table 7.7) so that it can be seen which items are easy and which are hard. The easiest item was on Topic 2 My Family Supporting Sentences and the hardest items was on Topic 1 My *Idul Fitri* Holiday Concluding Sentence.

Pretest

The pretest Writing data with 12 ‘items’ were also analysed with the RUMM2030 computer program, in a similar way to posttest Writing data. The results were similar for both pretest and posttest data and, to avoid repetition, the pretest Writing data output is not reported here. The Person Separation Reliability (based on

Rasch parameters) for the pretest Writing data was 0.91 and satisfactory. The Cronbach Alpha reliability for the pretest Writing data (based on the raw marking scores) for the pretest Writing data was 0.93 and also satisfactory.

Discussion and Summary

This chapter shows the results of the data analysis for the Writing test as part of the process of teaching and learning English as a second language in Aceh Province. Out of the original 18 'items', six items showed misfit to the Rasch measurement model and were deleted, leaving 12 'items' to create a linear measure. The Fit Residual data was partially satisfactory, showing some reasonable item-person response patterns, but it was not ideal. The Item-Trait Interaction and principal components analysis of the residuals indicated that two factors were present in the data, that is, the eigenvalue was 2.9, and was therefore above the chance level. However, the Person Separation Index of Reliability and the Cronbach Alpha Reliability were very satisfactory, indicating that there was good separation of measures in comparison to errors. The threshold values and the Scoring Category Curves showed that the scoring categories were used consistently and logically. The Item Characteristic Curves showed good discrimination. So a reasonable linear scale was created which was then used for the ANOVA and ANCOVA analyses which are presented in the next chapter.

The Item Characteristic Curves showed that none of the 12 items exhibited any statistically significant differential item functioning (DIF) by gender and by type of language of instruction (bilingual and monolingual). However, over the Rasch measure for all 12 items together, bilinguals had statistically significantly better Writing ability than monolinguals and girls had statistically significantly better Writing measures than boys, the latter being true even in the pretest, Monolingual Writing data.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for English Writing – based on the linear Rasch-created measures.

CHAPTER EIGHT

DATA ANALYSIS (PART 4)

EXPERIMENTAL RESULTS FOR ENGLISH WRITING

This chapter presents a comparison between the linear Rasch measures for the Writing Test for the bilingually-taught group, also known as the experimental group, and similar measures for the monolingually-taught group, also known as the control group, in a pretest/posttest, control/experimental group design. The measures on the pretest have been equated with those on the posttest in order that the Writing Test measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group are all on the same scale and thus comparable.

A Mixed Design ANOVA, also known as the Split-Plot ANOVA, was employed with the computer program SPSS (Statistical Package for the Social Sciences, Pallant, 2007) to compare the measures of Writing between the two language programs (bilingual program and monolingual program) and for gender (female and male students). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Writing posttest, administered following the completion of the two months of teaching English as a foreign language. The Writing pretest was conducted prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences. Tests for main and interaction effects using the pretest measures as the covariate were used. In order to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity (equality of variances of differences between all combinations of groups), preliminary checks were conducted and found to be satisfactory.

Results for Mixed Design ANOVA (Split-Plot Design ANOVA)

The results showed that there was a significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 0.818, $F(1,778) = 173.41$, $p = 0.000$, partial eta squared (η_p^2) = 0.182. According to Cohen's effect size guidelines, 0.921 is a large effect, as well as 0.481 (Cohen's η_p^2 values: 0.14 = large effect; 0.06 = medium effect; and 0.01 = small effect) (Cohen, 1988; also cited in Burns, 2000). Similarly, there was a significant main effect for time (pretest and posttest), with Wilks' Lambda = 0.79, $F(1,778) = 9089.87$, $p = 0.000$, $\eta_p^2 = 0.921$ for the Writing Test. The main effect comparing the two types of intervention (bilinguals and monolinguals) was also significant, with $F(1,778) = 720.70$, $p = 0.000$, $\eta_p^2 = 0.481$. This suggested a real difference in students' English Writing where bilingual students were able to write English Writing better than monolingual students (see Figure 8.1 and Table 8.1).

It is interesting to note that the posttest measures are lower, on average, than the pretest measures. One explanation for this is that the two month time interval between pretest and posttest measures is insufficient for young Indonesian students to strongly improve their English writing skills. Another explanation is that there was an experimental student effect at pretest but not at posttest. It is possible that a combination of these two effects is responsible for the results.

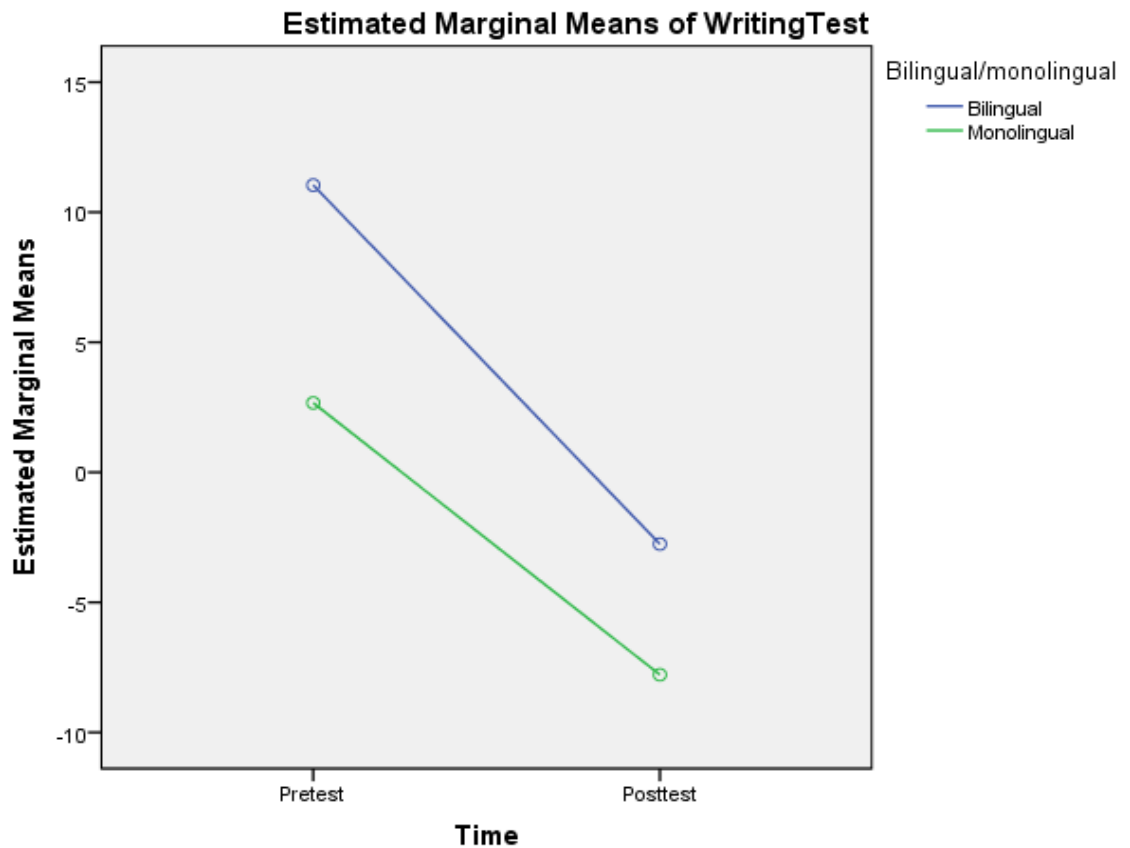


Figure 8.1 Graph of Writing Test by Group

Table 8.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	11.04	5.88	386	2.67	3.53
Posttest	394	-2.76	3.12	386	-7.78	2.04

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.991, $F(1,778) = 7.378$, $p = 0.007$, $\eta_p^2 = 0.009$ (a very small effect according Cohen's 1988 rules). On the other hand, the main effect for time (pretest and posttest) was significant, with Wilks' Lambda = 0.95, $F(1,778) = 7393.32$, $p = 0.000$,

$\eta_p^2 = 0.905$ (a large effect size according to Cohen, 1988) for the Writing Test.

Likewise, the main effect comparing the two types of intervention (monolingually-taught and bilingually-taught) was significant, with $F(1,778) = 47.614, p = 0.000, \eta_p^2 = 0.58$ (a medium effect size according to Cohen's 1988 regulations). This suggested that there was no real distinction in students' ability, between females and males (see Figure 8.2 for plot and Table 8.2 for pretest and posttest measures).

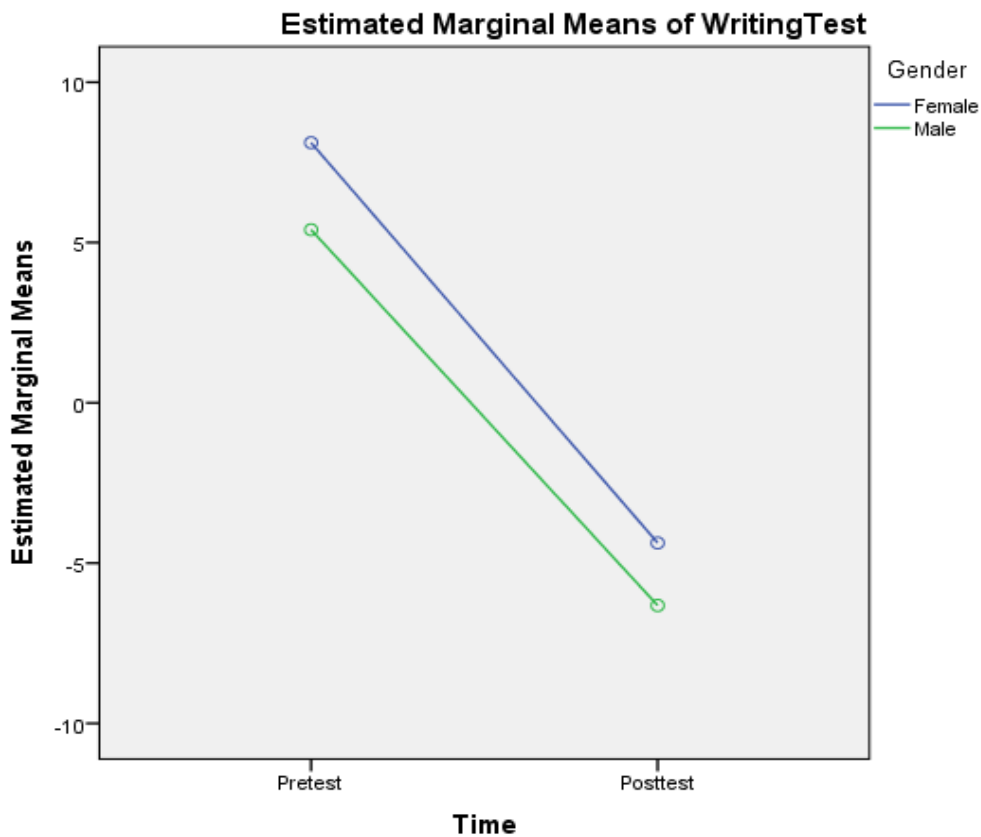


Figure 8.2 Graph of Writing Test by Gender

Table 8.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	8.11	6.72	349	5.40	5.68
Posttest	431	-4.37	3.65	349	-6.32	3.35

Figures 8.1 and 8.2, and Tables 8.1 and Table 8.2 indicate that girls have performed slightly better than boys, but not by much. This means that the difference was not significant.

Results for Two-Way ANCOVA

The Two-Way ANCOVA results for the Writing Test are given as follows. This was administered to assess the comparison of two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught), and gender (male and female students). The dependent variable was scores on the posttest, administered after the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language taking place (Time 2). Scores on the pretest administered prior to the start of the programs (Time 1) were used as a covariate to control for individual diversity.

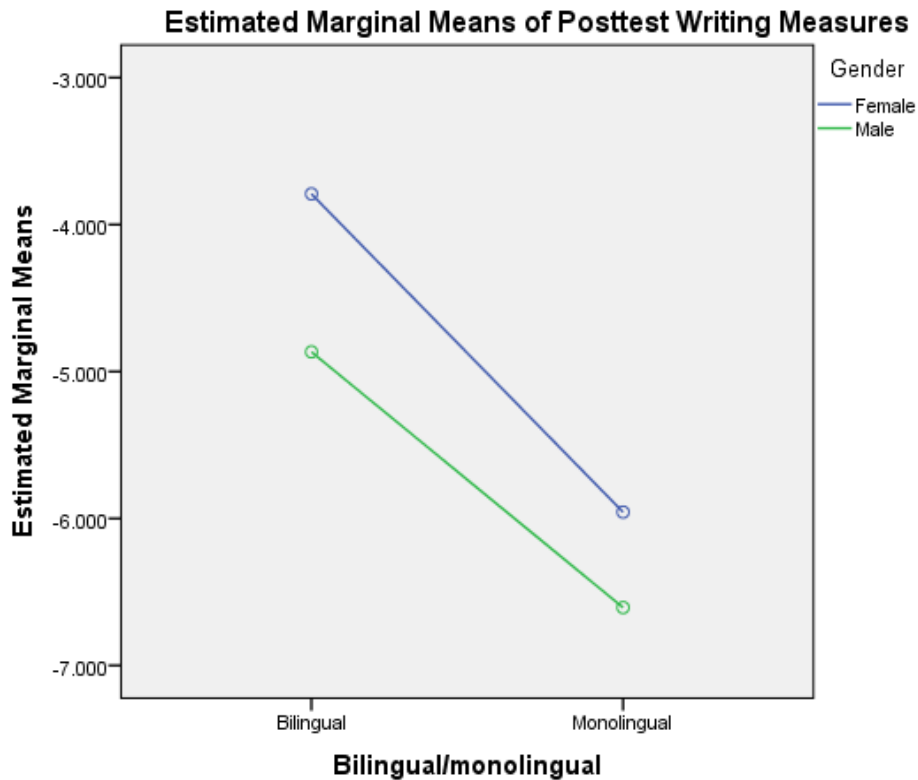
Table 8.3
Interaction Effect and Main Effect of Writing between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest Writing Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared (η_p^2)
Corrected Model	7728.990 ^a	4	1932.248	571.979	.000	.747
Intercept	15459.092	1	15459.092	4576.162	.000	.855
Pretest RC Rasch Measure on Posttest Scale						
Bilingual monolingual	421.216	1	421.216	124.687	.000	.139
Gender	135.485	1	135.485	40.106	.000	.049
Bilingual monolingual * Gender						
Gender	8.632	1	8.632	2.555	.110	.003
Error	2618.089	775	3.378			
Total	31791.159	780				
Corrected Total	10347.079	779				

a. R Squared = .747 (Adjusted R Squared = .746)

In order to confirm that there were no violations of the assumption of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable

measurement of the covariate, preliminary checks were carried out and found to be satisfactory. After adjusting for the pretest Writing measures, the result confirmed that there was no significant interaction effect between the type of programs (bilinguals and monolinguals): $F(1, 775) = 2.555, p > .005 (.110)$, with a small effect size (partial eta squared = 0.003) (see Table 8.3 for pretest-posttest measures, and Figure 8.3 for plot).

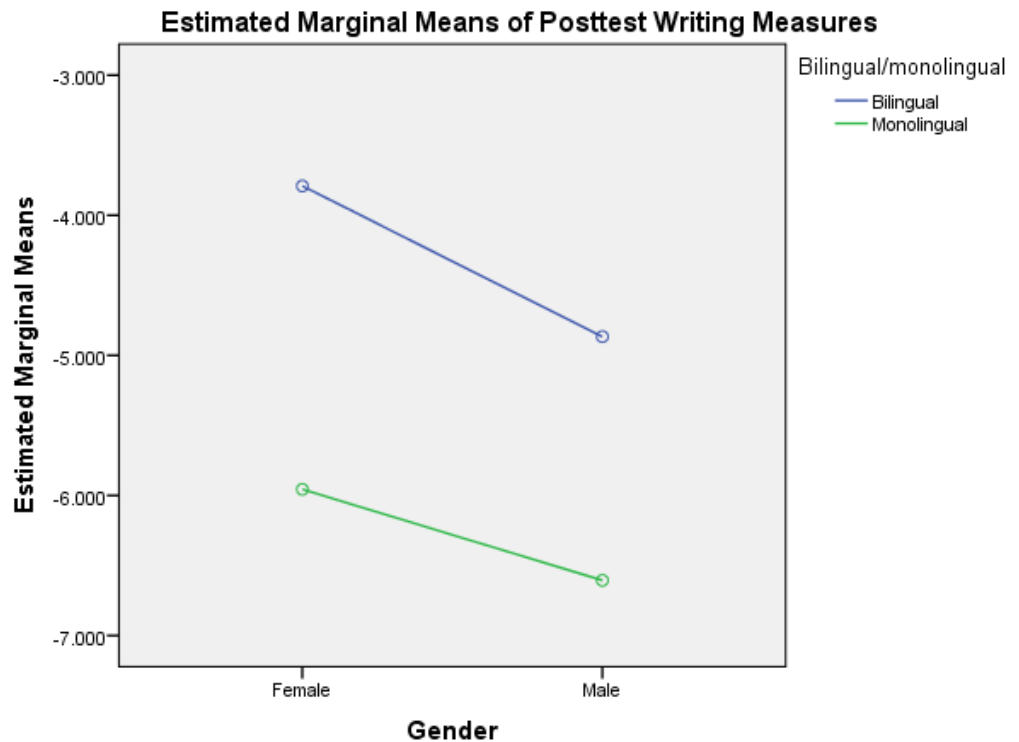


Covariates appearing in the model are evaluated at the following values: Pretest Writing Rasch Measure on the Posttest Scale = 6.90

Figure 8.3 Graph of Writing Test by Group (Bilingual/Monolingual) with Pretest as Covariate

On the other hand, the main effect for the language program was statistically significant with $F(1, 775) = 124.69, p = 0.000$, with $(\eta_p^2) = 0.139$ (considered a large effect size according Cohen’s 1988 regulations); as well as for gender, with $F(1, 775) = 40.106, p = 0.000$ with $(\eta_p^2) = 0.49$ (a very large effect size) (see Figure 8.4 for plots).

These results disclose that female students' score on the Writing Test was better than those of male students; and female and male students in the bilingual program responded better than female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest Writing Rasch Measure on the Posttest Scale = 6.90

Figure 8.4 Graph of Writing Test by Gender with Pretest as Covariate

Summary

The same English Writing pretest and posttest were given to bilinguals (N = 394) and monolinguals (N = 386) within a two month period (equivalent to 16 meetings), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA with the SPSS computer program and the pretests were used as the covariate. There are two main sets of findings: (1) with regards to the Mixed Design, and (2) with regards to the Two-Way ANCOVA.

The main findings, with regards to the Mixed Design ANOVA, are:

1. There was a significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 0.818, $F(1,778) = 173.41$, $p=0.000$, partial eta squared = 0.182 (big effect).
2. There was a main effect for time (pretest and posttest), with Wilks' Lambda = 0.79, $F(1,778) = 9089.87$, $p = 0.000$, partial eta squared = 0.921 (very large effect).
3. The main effect comparing the two types of intervention was significant, with $F(1,778) = 720.70$, $p = 0.000$, partial eta squared = 0.481 (large effect).
4. There was no statistically significant interaction between gender and time, with Wilks' Lambda = 0.991, $F(1,778) = 7.378$, $p = 0.007$, partial eta squared = 0.009 (small effect).
5. The main effect for time (pretest and posttest) was significant, with Wilks' Lambda = 0.95, $F(1,778) = 7393.32$, $p = 0.000$, partial eta squared = 0.905 (very large effect).
6. The main effect comparing the two types of intervention was significant, with $F(1,778) = 47.614$, $p = 0.000$, partial eta squared = 0.58 (large effect).

The main findings, with regards to the Two-Way ANCOVA, are

1. There was no significant interaction effect between the type of programs (bilinguals and monolinguals): $F(1, 775) = 2.555$, $p > .005$ (.110), with partial eta squared = 0.003 (small effect).
2. The main effect for the language program was statistically significant, with $F(1,775) = 124.69$, $p = 0.000$, partial eta squared = 0.139 (small effect).
3. The main effect for the gender was significant, with $F(1,775) = 40.106$, $p=0,000$ with partial eta squared = 0.49 (large effect).

The results revealed that bilingual students were able to write English essay better than monolingual students. Even though there was no real distinction in students' ability on English Writing between females and males, female students from both

bilingual and monolingual schools scored better on the writing test than male students from bilingual and monolingual schools. The posttest results for English Writing were measured to be lower, after two months, than the pretest results, and this was true for both females and males, and for both bilinguals and monolinguals. This was probably because young Indonesian students find writing in English to be hard and the English Writing test was marked hard (at a high level). In other words, it requires longer than two months for significant improvements in English Writing to occur.

The next chapter explains the Rasch analysis with the RUMM2030 computer program for the Attitude/Behaviour Questionnaire.

CHAPTER NINE

DATA ANALYSIS (PART 5)

RASCH MEASUREMENT OF ATTITUDE/BEHAVIOUR

This chapter presents the findings of the Rasch analysis of the Behaviour posttest data.

The original Attitude/Behaviour Questionnaire (I= 42 items) was designed to collect data which, when analysed with the Extended Logistic Model of Rasch (Andrich, 1988a), could be used to create a linear measure of students' English Behaviour at the posttest stage of the experiment. The items were answered in three response categories; 'never or rarely' (scored 0), 'some of the time' (scored 1), and 'most or all the time' (scored 2). The data were analysed with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010) to create a unidimensional linear scale.

Initial Analysis

The analysis involved running the data through the RUMM2030 program and checking for non-fitting items and items with poor discrimination, which were then deleted, and re-running the program. After four analyses, all the 21 attitude items (except item 38) were deleted because of misfit to the measurement model. This was not consistent with the model used to develop the Behaviour Questionnaire that was based on many previous studies where attitude and behaviour were measured together (see Waugh, 2003, 2005, 2010a, 2010b). The RUMM program does not tell the researcher why an item doesn't fit the Rasch Measurement Model, just that it doesn't fit. It was difficult to see why the attitude items didn't fit the measurement model, but the students were in their first year of being bilingually-taught and, because they were not strong in English reading, the misfit may have been primarily due to their low command of reading and understanding of English, related to the different classroom culture in the two groups, in a test situation. Thus, it may be that the two groups of students (bilingually-taught and monolinguals) did not have agreement on the difficulties of the items because of their differences in their command of English, combined with some

difference in culture (living in an Indonesian culture, and learning English culture through English lessons), and that was the substantial cause of the misfit. The fifth analysis, consisting of ten behaviour items (items 2,8,10,12,16,18,26, 28, 36 and 38) and one attitude item (item 7), produced a good fit to the measurement model. Deletion of the attitude item 7 (“*I say new words several times in English*”) and a re-analysis with the ten behaviour items produced a worse fit to the measurement model and so the attitude item 7 was re-instated. The following material shows the output from the RUMM program when a good, unidimensional, linear scale of Behaviour with respect to Learning English was created with 11 items.

Output from Final Analysis

Standardised Fit Residuals

The Fit Residual data for both items and students have a mean near zero and a standard deviation near one when the data fit the measurement model. The mean for items is 0.195, and for persons is -0.306. The Standard Deviation is 1.112 for items and 1.525 for persons. It shows that the data fit the measurement model satisfactorily and this means that there is a good consistency of item-student response pattern. Table 9.1 shows item and person fit to the measurement model for the Behaviour measure.

Table 9.1
Overall Fit Statistics for the Behaviour Measure (N=779, I=11)

	ITEM-PERSON INTERACTION			
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	0.195	0.458	-0.306
Standard Deviation	0.217	1.112	0.972	1.525

Notes:

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model.

Dimensionality and Item-Trait Interaction

The item trait interaction chi-square was 103.82, df=99, and p=0.35 (see Table 9.2). This indicated that there was good agreement amongst the students about the item difficulties all along the scale. This, in turn, means that a single parameter for each

student, the person measure, and a single parameter for each item, the item difficulty, can be used to accurately predict each student's response to each item. A principal components analysis of the residuals showed that the first eigenvalue was 1.45, which is within the chance value, supporting the view that a unidimensional measure has been made.

Table 9.2
Item-Trait Interaction for Behaviour Scale

Total Item Chi-Square	103.82
Separation Index	0.71
Total Degree of Freedom	99.00
Total Chi-Square Probability	0.35
Cronbach Alpha	0.73

Notes:

1. The Index of Person Separation (Andrich & Van-Schoubroeck, 1989) is interpreted like a Cronbach Alpha (Cronbach, 1951) and is good.
2. The item-trait interaction test is a chi-square. The results indicate that there was a good collective agreement amongst the students about the item difficulties along the scale.
3. All numbers are given to two decimal points because the errors are only up to two decimal points.

Person Separation Index

For a good measure, a Person Separation Index is desired to be 0.75 or greater, as it is an indicator that the student measures are separated by more than their standard error. The Person Separation Index for this study is 0.71 (see Table 9.2 above). It indicates that there is a reasonable separation of measures in relation to the error (which is about 0.08 logits). While the Person Separation Index is calculated on the Rasch-created parameters, the Cronbach Alpha (Cronbach, 1951) is calculated on the raw scores. In this case, the Cronbach Alpha was 0.73, indicating reasonable reliability of the data.

Individual Item Fit

The RUMM 2030 program calculates individual item fits to the measurement model. For the Attitude and Behaviour measure, all 11 items fit the measurement model (see Table 9.3). In addition, all the standardized residuals fall within the range -1.2 to +1.1, supporting the good fit to the measurement model (which usually has to be within plus or minus 2 SDs).

Table 9.3*Locations, Standard Errors, Residuals and Chi-Squares for Behaviour Items*

Item No.	Location	SE	Residual	DF	Chi-Square	DF	Probability
2	-0.270	0.069	0.471	706.27	5.065	9	0.83
7	-0.173	0.056	0.946	706.27	8.522	9	0.48
8	0.133	0.060	-0.098	706.27	4.401	9	0.88
10	0.129	0.059	0.699	706.27	15.060	9	0.01
12	0.382	0.059	0.030	706.27	7.224	9	0.61
16	-0.281	0.059	-1.133	706.27	14.555	9	0.10
18	0.101	0.058	0.314	706.27	5.077	9	0.83
26	0.065	0.059	0.229	706.27	6.843	9	0.65
28	0.185	0.057	0.563	706.27	4.026	9	0.91
36	-0.481	0.059	-0.225	706.27	11.377	9	0.25
38	0.211	0.057	1.072	706.27	9.999	9	0.35

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.
3. DF means degrees of freedom. Probability is based on the chi-square fit to the measurement model.

Threshold Values

Items thresholds are positions on the scale between adjacent response categories where the odds are 1:1 that students will respond to a particular item, in either category. It is expected that the students would use the thresholds in the way that they were intended by the researcher and so the thresholds should be ordered in line with the conceptual ordering of the response and scoring categories. The thresholds in this measure were ordered in line with the conceptual ordering of the response and scoring categories (see Table 9.4).

Table 9.4
Item Thresholds Uncentralised (Item=11, Number=779) for Behaviour Measure

Item	Item Location	Thresholds	
		1	2
2	-.222	-1.805	1.361
7	-.132	-.804	.539
8	.157	-1.188	1.504
10	.124	-1.038	1.287
12	.352	-.623	1.327
16	-.142	-1.144	.858
18	.061	-.935	1.058
26	.092	-.975	1.160
28	.030	-.831	.893
36	-.439	-1.109	.230
38	.117	-.831	1.066

Note: The thresholds are ordered in line with the scoring categories.

Scoring Category Curve

The RUMM 2030 program produces curves of the scoring categories for each item. The Scoring Category Curves show the relationship between the probability of scoring in each category. Each item has three response categories: ‘Never or rarely’ (scored 0); ‘Some of the time’ (scored 1); and ‘Most or all the time’ (scored 2). The Scoring Category Curves should show a consistent relationship between the probability of scoring and the measure from low to high indicating that the scoring was done consistently and logically. A Scoring Category Curve for Item 2 is shown in Figure 9.1.

Figure 9.1 (Item 2: “*I pay attention to someone speaking English*”) showed that the scoring was done logically and consistently. When students have low measures on item 2, then they have a high probability of obtaining a zero score (the lowest response); and when they have a medium measure, they have a medium probability of scoring 1 (the moderate response); and when they have a high measure, they have a high probability of scoring 2 (the highest response). The Scoring Category Curves for the other items were checked and they showed logical and consistent scoring as well.

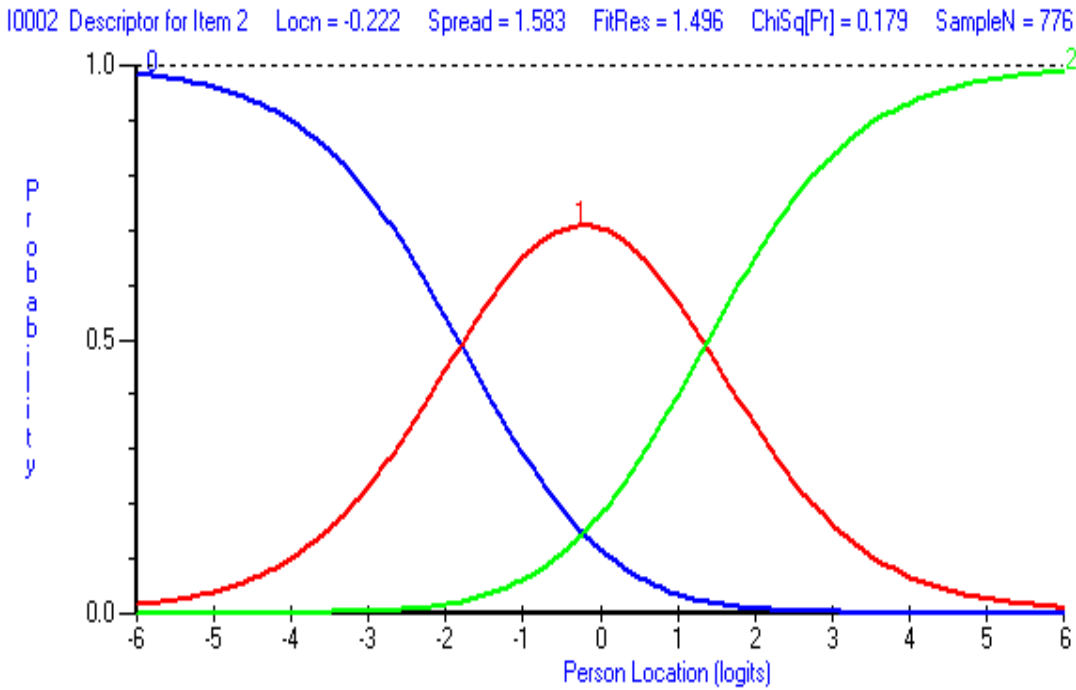


Figure 9.1 Item Category Curve for Item 2

Note: The blue curve is for a score of 0, the red curve for a score of 1 and the green curve for a score of 2.

Item Characteristic Curves

The Item Characteristic Curve provides information on item differentiation between persons and the item location. A group of students is considered to have performed well if their values (in the form of black dots) fit on the ogive curve. The ogive curve is the expected values for an item against the student measures (low to high). Take as an example Item 2 (Figure 9.2). With most of the dots close to the curve, it is considered a good fit to the measurement model and shows good discrimination. The Item Characteristic Curves for the other items were checked and found to be satisfactory.

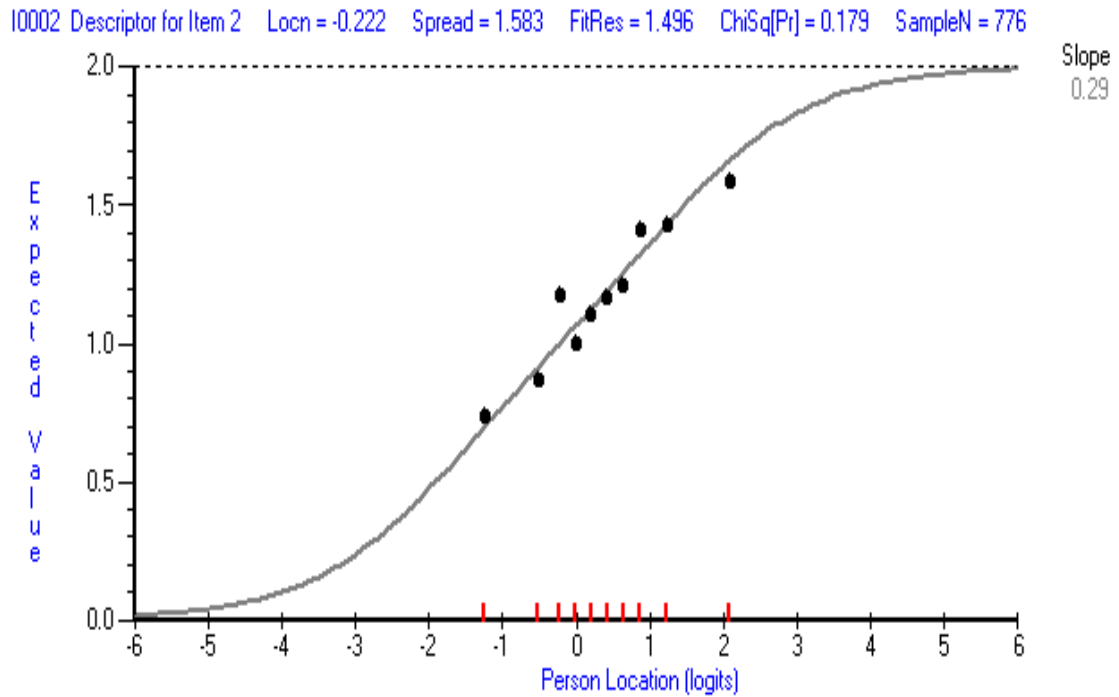


Figure 9.2 Item Characteristic Curve for Item 2

Person-Item Threshold Distribution (Targeting)

Measures are considered well-targeted when the range of item thresholds and the range of student measures are about the same. It means that the items are not too hard and not too easy for the students. Figure 9.3 shows the targeting graph for the measure of Behaviour. The targeting of the items is not as good as expected – there were insufficient easy, medium difficulty and hard items – but this is because many of the original items did not fit the measurement model and were deleted in the initial analysis. In the initial design of the items, there was a wide range of item difficulties but, apparently, the bilinguals and the monolinguals did not agree on the difficulties for most items, causing them to be deleted from this analysis.

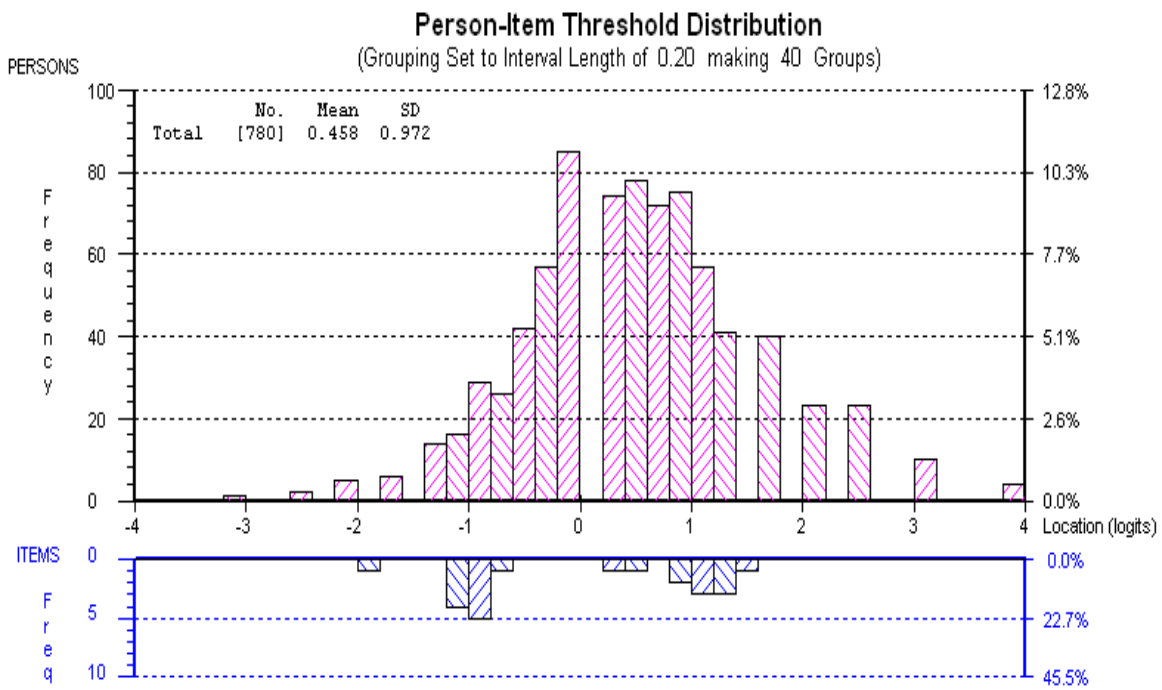


Figure 9.3 Person-Item Threshold Distribution for Behaviour

Figure 9.4 shows the distribution of item thresholds and the students on a ‘map’ and this also shows the restricted range of item thresholds. Somehow, this targeting problem would have to be rectified in any future use of the scale for these students.

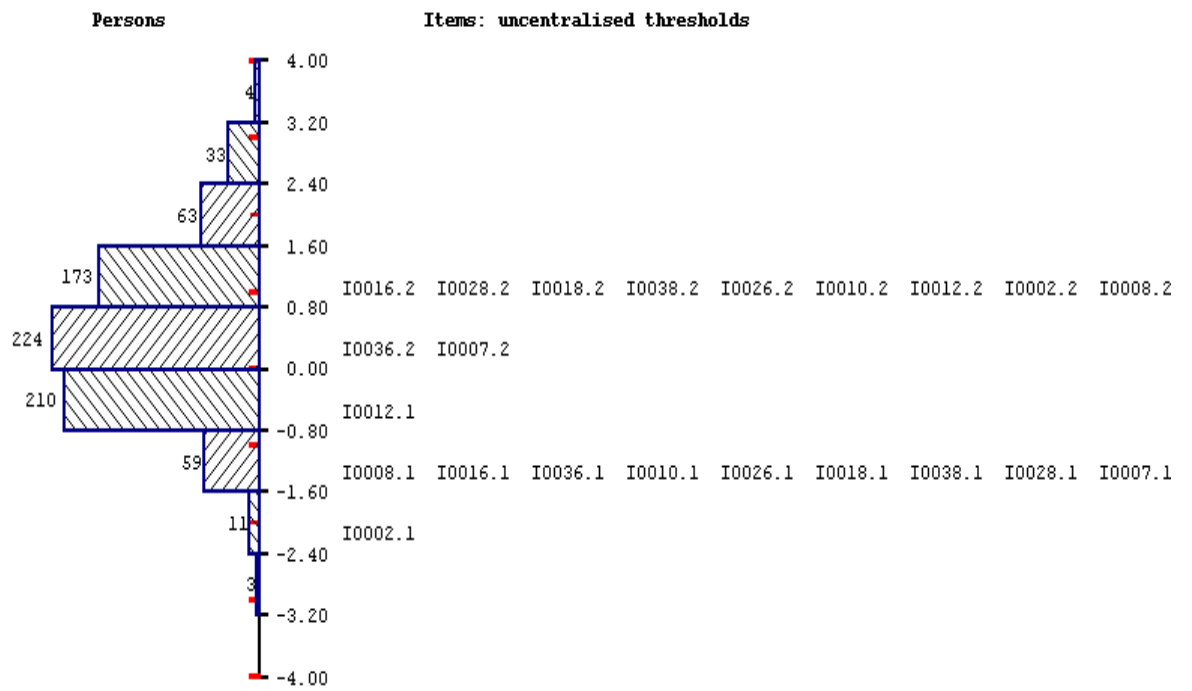


Figure 9.4 Behaviour Map

Note: I0016.2 means threshold 2 for item 16, I0002.1 means threshold 1 for item 2, and so on.

DIF by Gender

Each of the 11 items of the Behaviour measure showed no statistically significant Differential Item Functioning (DIF) by gender (see Figures 9.5, 9.6 and 9.7, for examples).

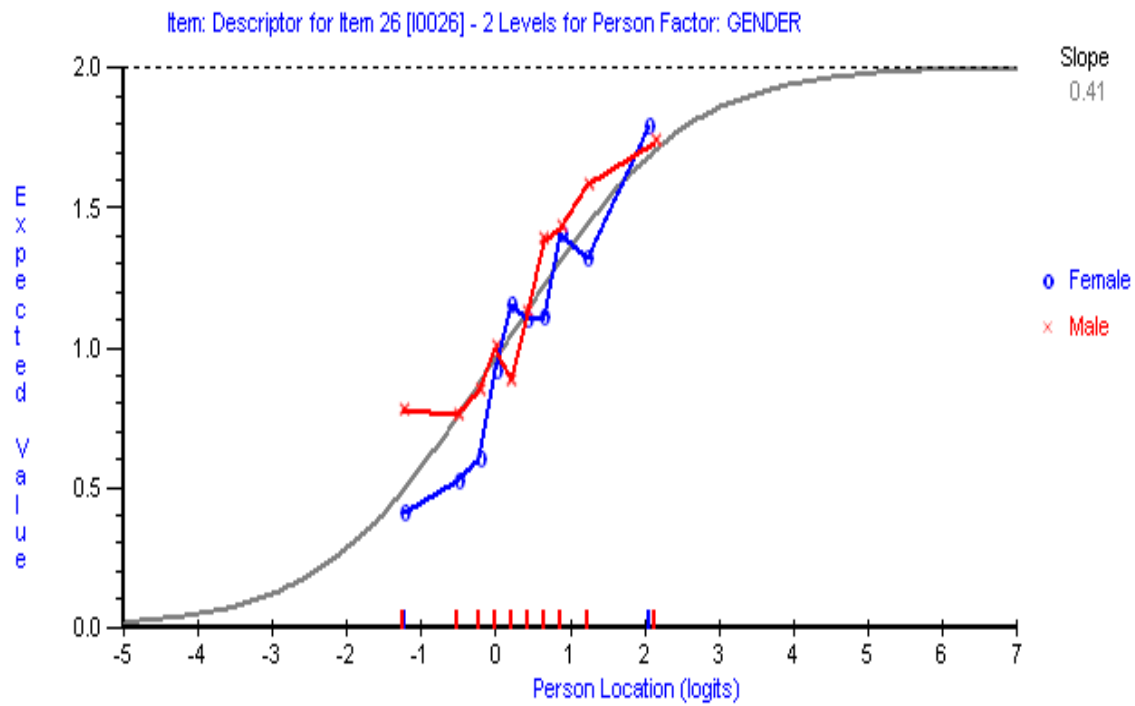


Figure 9.5 Item Characteristic Curves by Gender for Behaviour Item 26

Note: No statistically significant main effect by gender, $F=5.02$, $df=19,1$, $p=0.025$.

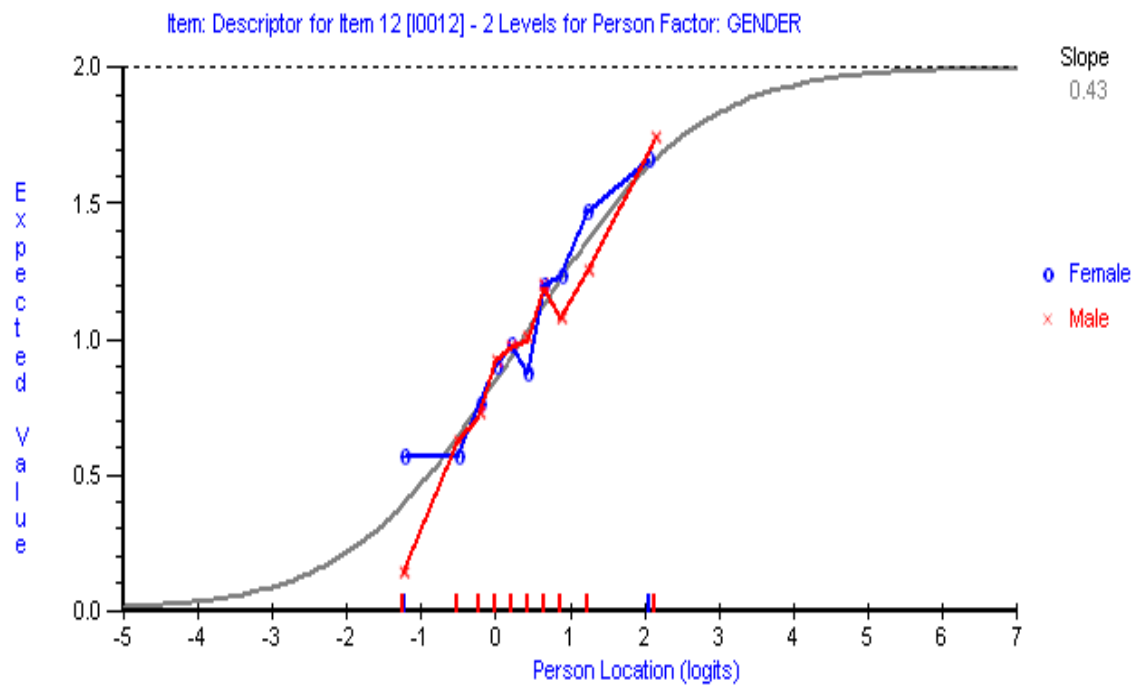


Figure 9.6 Item Characteristic Curves by Gender for Behaviour Item 12

Note: No Statistically significant main effect by gender, $F=2.96$, $df=19,1$, $p=0.09$.

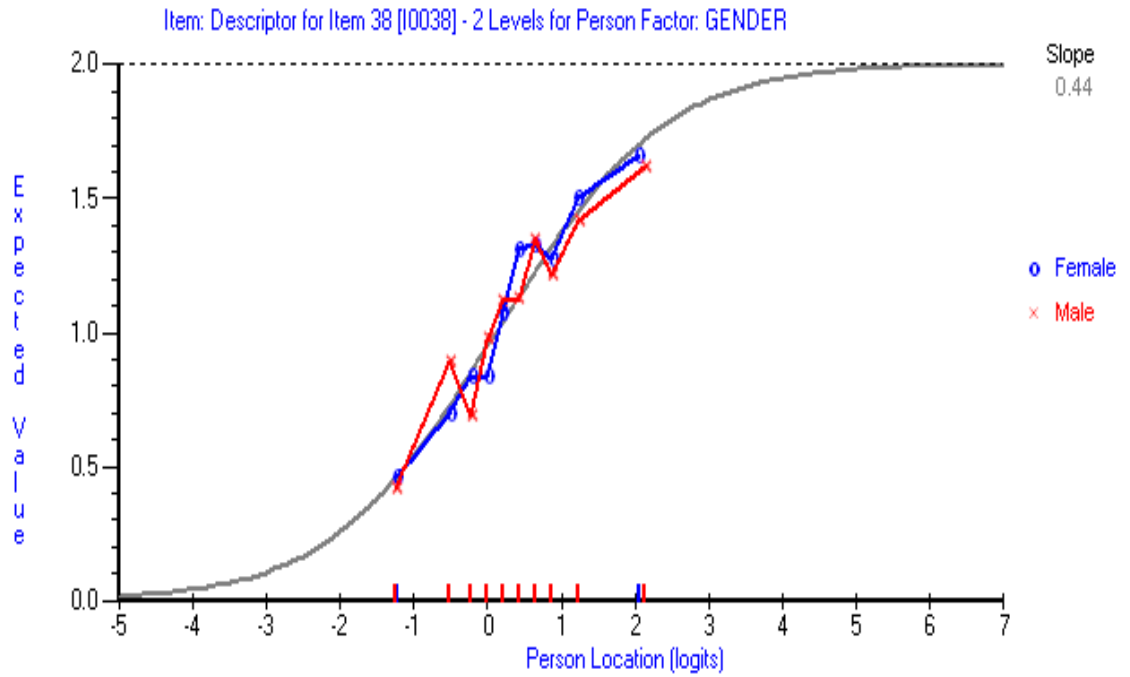


Figure 9.7 Item Characteristic Curves by Gender for Behaviour Item 38

Note: No statistically significant interaction effect by gender, $F=0.19$, $df=19,1$, $p=0.66$

All the Item Characteristic Curves for the other items were checked and showed no statistically significant difference by gender.

DIF by Type of Language Instruction

Only one item (Item 7) showed DIF by type where bilinguals had improved results on the Behaviour measure compared to monolinguals ($F=23.81$, $df=19,1$, $p=0.00000$, see Figure 9.8). Figures 9.9 and 9.10 show that there was no DIF for item 2 and item 38. However, over the Rasch measures for all the 11 items together, bilinguals had a statistically significantly higher Behaviour measure than monolinguals ($F=20.56$, $df=1,778$, $p=0.0000$, see Figure 9.12).

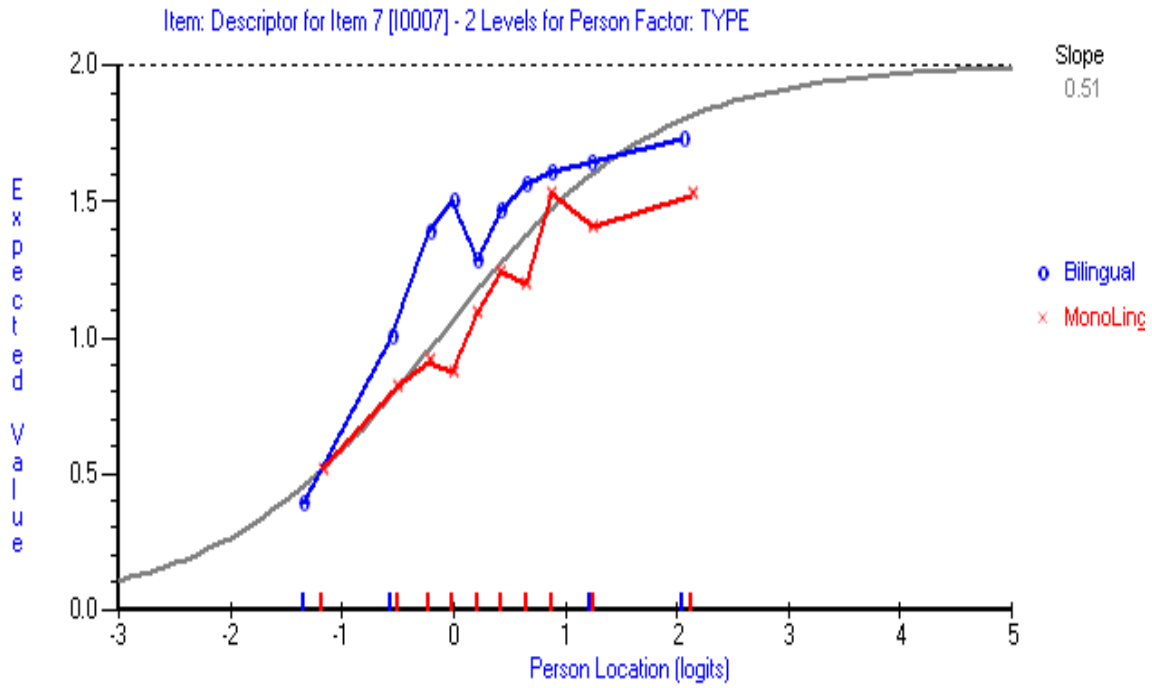


Figure 9.8 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 7

Note: There is a statistically significant main effect by type, $F=23.81$, $df=19,1$, $p=0.00000$

Bilinguals have a statistically significantly higher Behaviour measure.

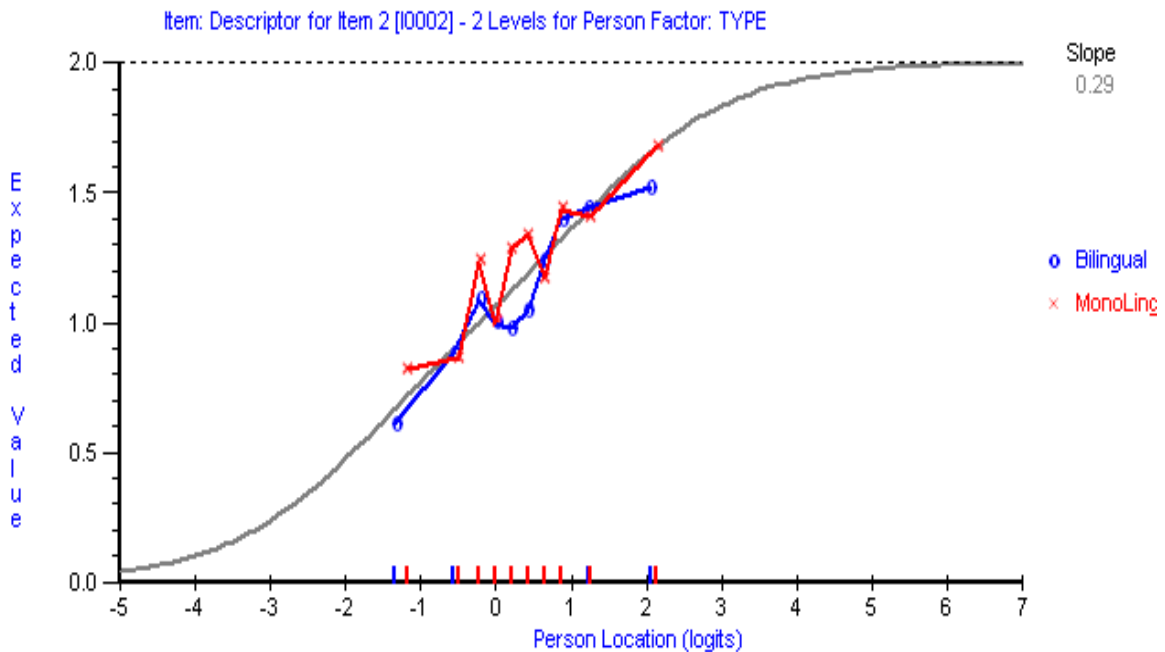


Figure 9.9 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 2

Note: Not statistically significant by type of teaching ($F=5.65$, $df=19,1$, $p=0.02$)

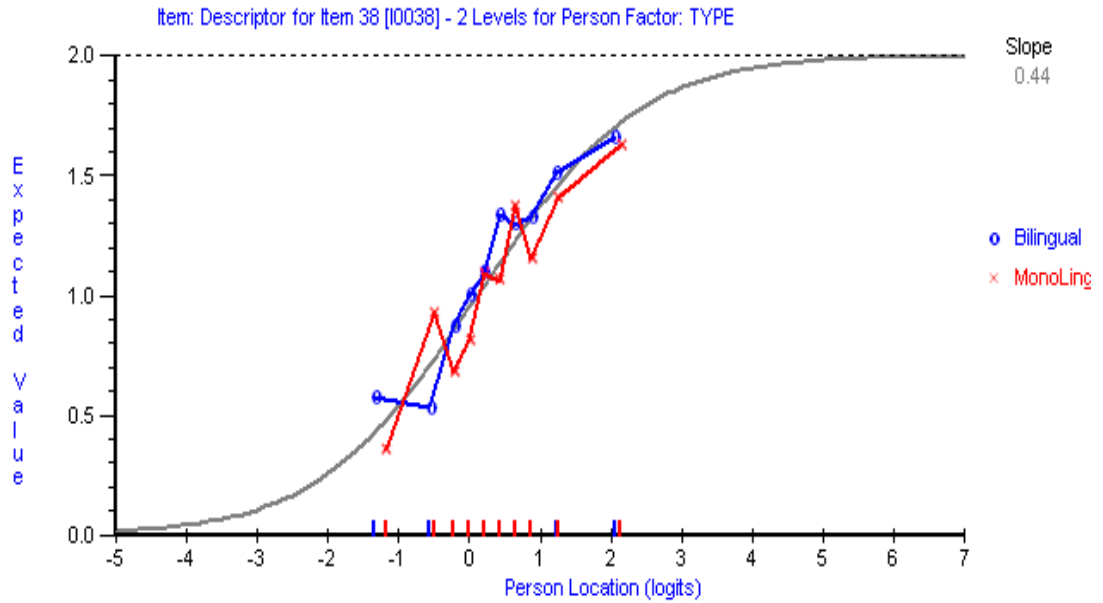


Figure 9.10 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 38

Note: Not statistically significant by type of teaching $F=4.66$, $df=19,1$, $p=0.03$

Targeting by Gender and Type of Language Instruction

Females showed a statistically significantly better Behaviour measure than males ($F= 11.52$, $df=1,778$, $p=0.0007$, see Figure 9.11) and bilinguals have a statistically significantly better Behaviour measure than monolinguals ($F= 20.56$, $df=1,778$, $p=0.00002$, see Figure 9.12).

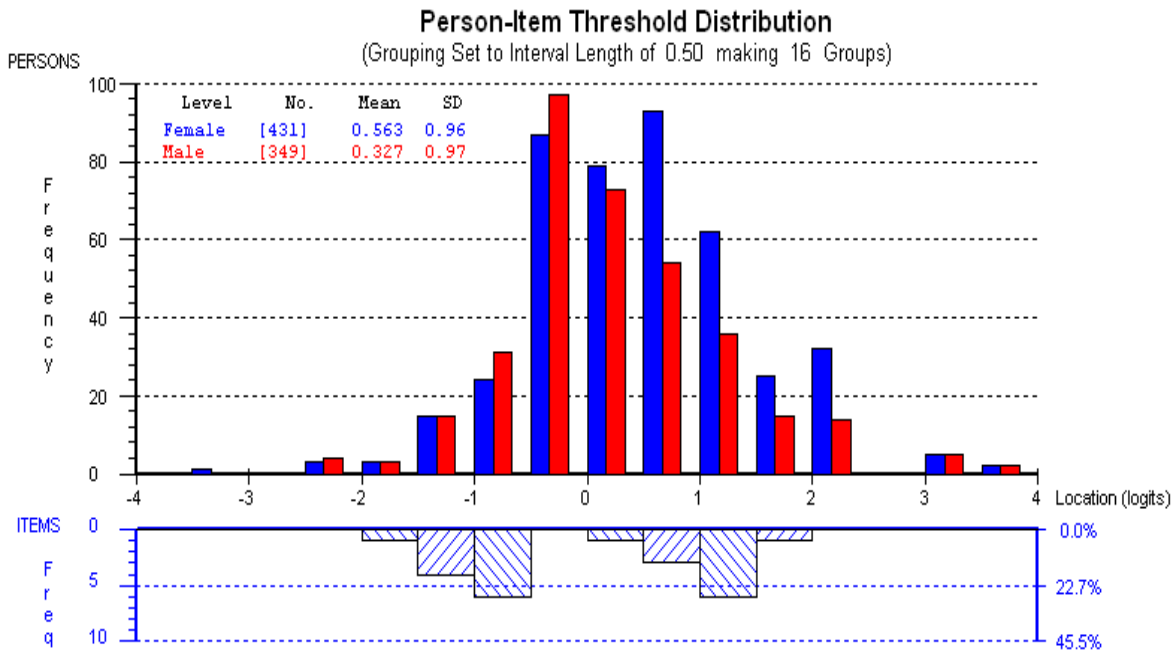


Figure 9.11 Targeting of Behaviour by Gender

Notes:

1. The person measures are on the upper-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the lower-side side from easy (LHS) to hard (RHS). $F= 11.52$, $df=1,778$, $p=0.0007$, which is a statistically significant difference.

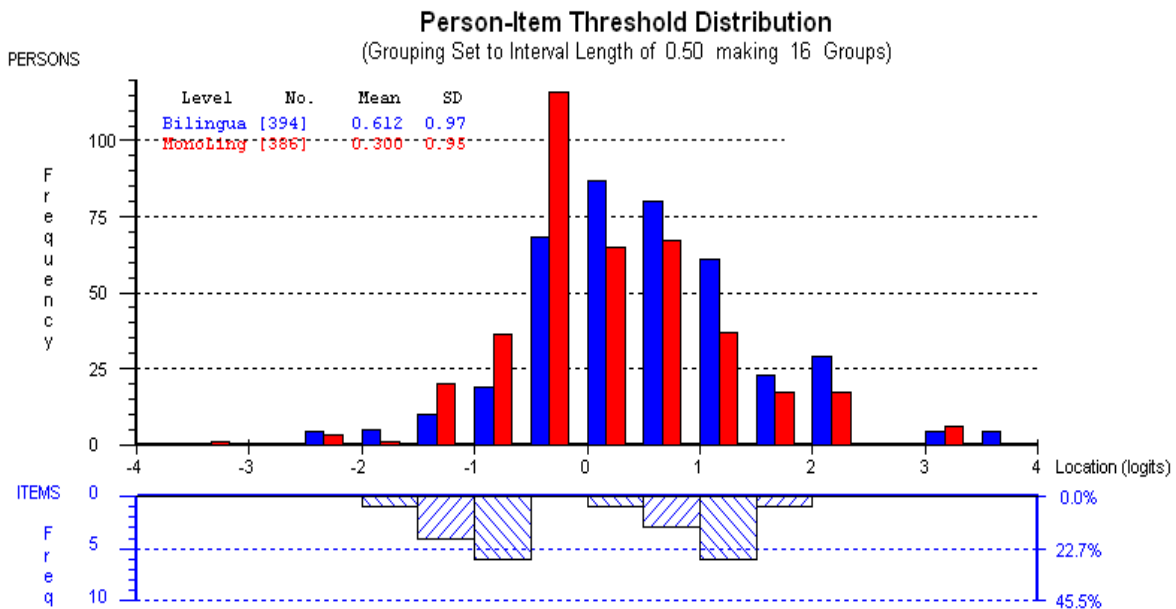


Figure 9.12 Targeting of Behaviour by Type of Language Instruction

Note: $F= 20.56$, $df=1,778$, $p=0.00002$ which is statistically significant for bilingual teaching.

Scale of Item Difficulties

Table 9.5 shows the item wording for Behaviour. The items have been ordered by difficulty from the easiest to the most difficult on the linear Rasch-created scale. The ordering of the items is consistent with the initial predicted conceptualized order, supporting the construct validity of the scale.

Table 9.5
Order of Difficulty of Items on the Linear Scale

Item No.	Item Location	Item Statements on the Behaviour
36 (easy)	-0.44	I actually like the way my teacher teaches English writing.
2	-0.22	I actually pay attention to someone speaking English.
16	-0.14	I actually read carefully words in English.
7	-0.31	I wish to say new English words several times.
28	0.01	I actually can understand English better when I do activities with friends.
18	0.06	I actually can read English at home on my own.
26	0.09	I actually learn more when I study English in groups.
38	0.12	I actually like English because we use it in the classroom.
10	0.13	I actually practice English with other students.
8	0.16	I actually say new English words several times.
12 (hard)	0.35	I actually start conversation in English with my friends.

Note: Item difficulties (locations) are measured in logits, the log odds of answering successfully.

The items were ordered from easy to hard on a linear scale (see Table 9.5) so that it can be seen which items are easy and which are hard. The easiest item involved the students' preference for the way their English teachers teach in the classroom, as expected (item 36 difficulty = -0.44 logits). The hardest item involved oral skill in ability to initiate speaking in English with their friends, as expected (item 12 difficulty = +0.35 logits).

Discussion and Summary

This chapter presents the results of the data analysis for the process of student behaviour in learning English as a second language in Aceh. Eleven items from the original 42 items produced a linear, unidimensional measure (31 items were deleted). The Fit Residual data showed that there was a good consistency for the item-person

response pattern. The Item-Trait Interaction (dimensionality) indicated that there was good agreement about the item difficulties along the scale. The Person Separation Index indicated that there was good separation of measures in comparison to errors. All items fitted the Rasch measurement model. The threshold values and the Scoring Category Curves showed that the scoring categories were used consistently and logically. The Item Characteristic Curves showed reasonable discrimination. All these data support the view that a linear, unidimensional measure of Behaviour was created so that valid inferences could be made.

However, 21 items, initially considered to be conceptually valid, had to be deleted because of misfit to the measurement model, apparently because the bilinguals and the monolinguals did not agree on the item difficulties. The deletion of these items caused a targeting problem, where the final scale had insufficient items across the whole difficulty range of the student measures. Thus this problem needs to be investigated further so that there are sufficient easy, medium and hard items to cover the full range of student measures in any future use of this scale.

The Item Characteristic Curves showed that the large majority of the items had no statistically significant differential item functioning (DIF) by gender and by type of language of instruction (bilingually-taught and monolingually-taught). However, over the Rasch measure for all 11 items together, bilinguals had statistically significantly better Behaviour measure than monolinguals and females had a statistically significantly better Behaviour measure than males.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for Reading Comprehension – based on the linear Rasch-created measures.

CHAPTER TEN

DATA ANALYSIS (PART 6)

EXPERIMENTAL RESULTS FOR BEHAVIOUR

In this chapter, the linear Rasch measures for Behaviour for the bilingually-taught group, known as the experimental group, were compared with similar measures for the monolingually-taught group, known as the control group, in a pretest/posttest, control/experimental group design. The measures on the pretest have been equated with those on the posttest, so that the Behaviour measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group are all on the same scale and are already comparable. The Mixed Design ANOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant, 2007) was conducted to access the comparisons between two different interventions (bilingually-taught program and monolingually-taught program), and to access effects of gender on scores/responses from the Behaviour Questionnaire at both the pretest and posttest. The Two-Way ANCOVA with SPSS was run to check if the independent variable interacted to predict the dependent variables. It was used to test for main and interaction effects with SPSS to compared bilinguals with monolinguals and females with males.

Results for Mixed Design ANOVA (Split-Plot Design ANOVA)

The Mixed Design ANOVA was conducted to compare the measure of Behaviour Questionnaire for the two language programs (bilingual program and monolingual program) and of gender (female and male). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Behaviour Questionnaire posttest, administered following completion of the two months teaching English as a foreign language. The Behaviour Questionnaire pretest was administered prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences.

A mixed between-within subjects analysis of variance (Mixed Design ANOVA or Split-Plot ANOVA) was conducted to assess the difference between the two interventions (bilingually-taught program and monolingually-taught program); and the comparison of gender, on participants' Behaviour measures, across two time periods (pretest and posttest).

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity. There was no significant interaction between the type of language program (Bilingually-taught program and Monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.27$, $p = 0.607$, partial eta squared = 0.00. Similarly, there was no main effect for time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.51$, $p = 0.697$, $(\eta_p^2) = 0.00$ for the Behaviour measure. However, the main effect comparing the two types of intervention was significant, with $F(1,778) = 29.131$, $p < 0.0005$, $(\eta_p^2) = 0.36$. According to Cohen's effect size guidelines, this main effect is large. This suggested a real difference in students' pretest and posttest in which bilingual students performed better in Behaviour measures than monolingual students (see Figure 10.2). It is noted that bilingual students started with a better responses on Behaviour measure pretest than monolingual students.

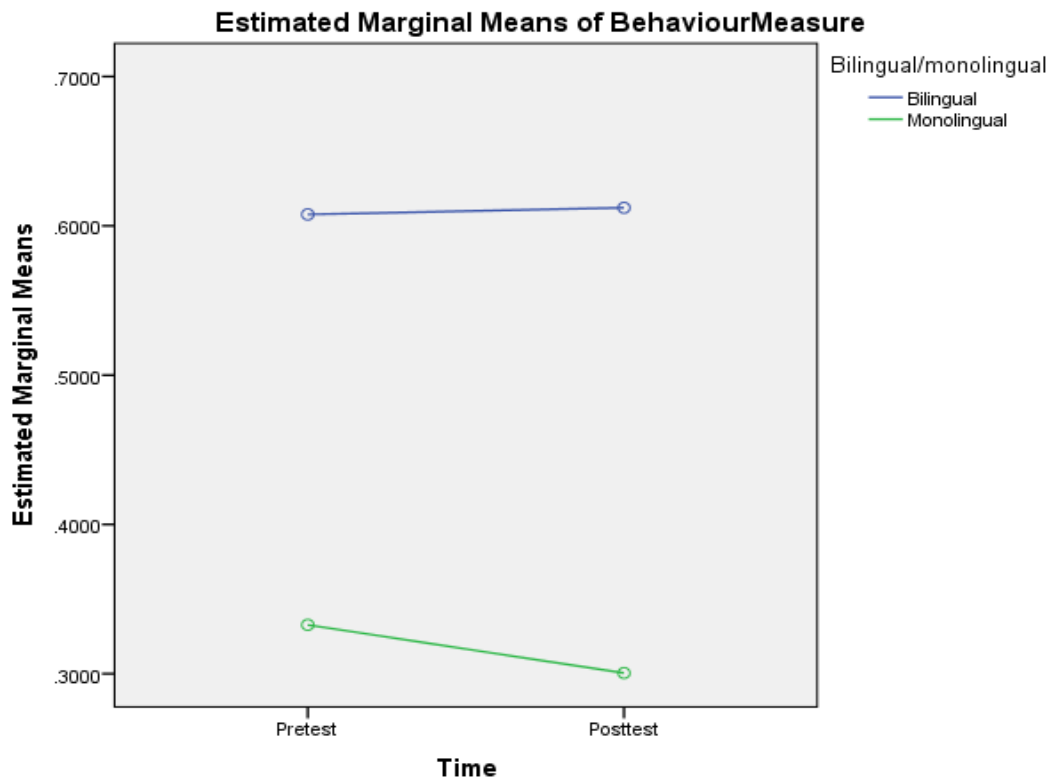


Figure 10.1 Graph of Behaviour by Group

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.996, $F(1,778) = 2.909$, $p = 0.89$, partial eta squared = 0.004 (a very small effect according Cohen's 1988 rules). Similarly, there was no main effect for time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.41$, $p = 0.839$, partial eta squared = 0.000 for the Behaviour measure. However, the main effect comparing the two types of intervention was significant, with $F(1,778) = 29.580$, $p < 0.0005$, partial eta squared = 0.37 (a large effect size according Cohen's 1988 regulations). This suggested a real difference in students' Behaviour measure, favouring females (see Figure 10.2 for plot).

Table 10.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	.607	.874	386	.332	.831
Posttest	394	.612	.970	386	.300	.948

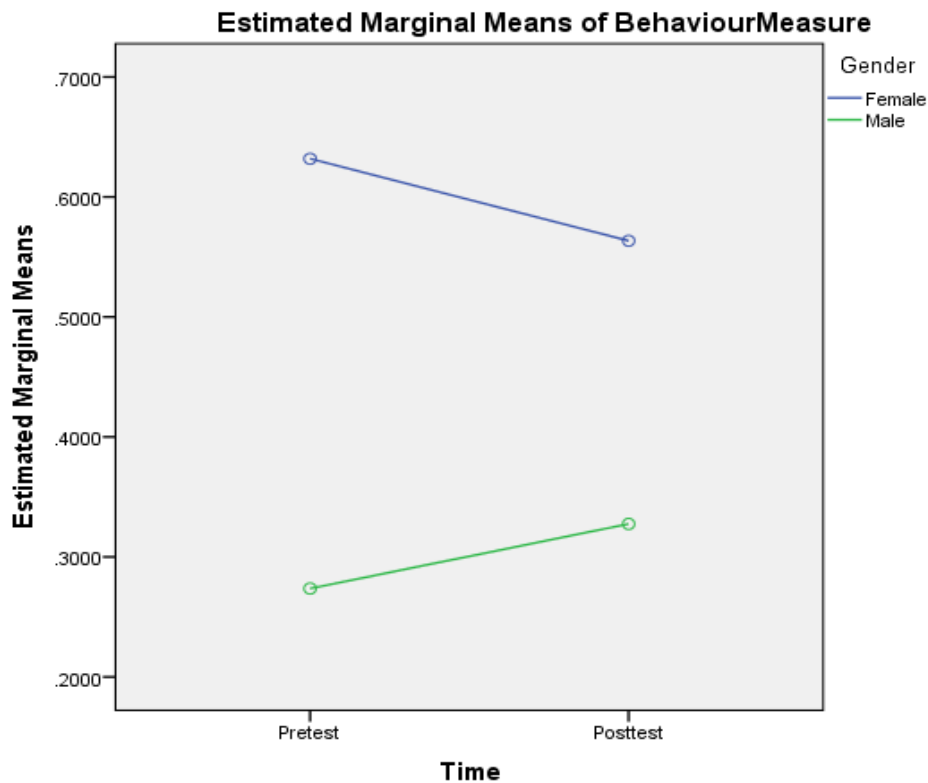


Figure 10.2 Graph of Behaviour by Gender

Table 10.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	.63	.85	349	.27	.84
Posttest	431	.56	.96	349	.33	.97

Figures 10.1 and 10.2, and Table 10.1 and Table 10.2 suggest that females have performed better than males. It is well-known that the English achievement of girls is better than that of boys (for example, see Machin & McNally, 2005), but are girls' behaviour with regards to learning English better than boys, as well? The findings from this study suggest that it is. It may be the better behaviour of girls with regard to learning English that is important and that, at least partially, explains the well-known finding that the English achievement of girls is better than that of boys.

Results for Two-Way ANCOVA

The following are the Two-Way ANCOVA results for the Behaviour measures. This was conducted to compare the two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught) and gender (male and female). The dependent variable was scores on the posttest, administered following the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language (Time 2). Scores on the pretest administered prior to the commencement of the programs (Time 1) were used as a covariate to control for individual differences.

Preliminary checks were conducted to ensure that there were no violations of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for the pretest of the Behaviour measure, there was no significant interaction effect: $F(1, 775) = .104, p > 0.05 (.747)$, with a small effect size ($\eta_p^2 = 0.000$) (see Table 10. 3).

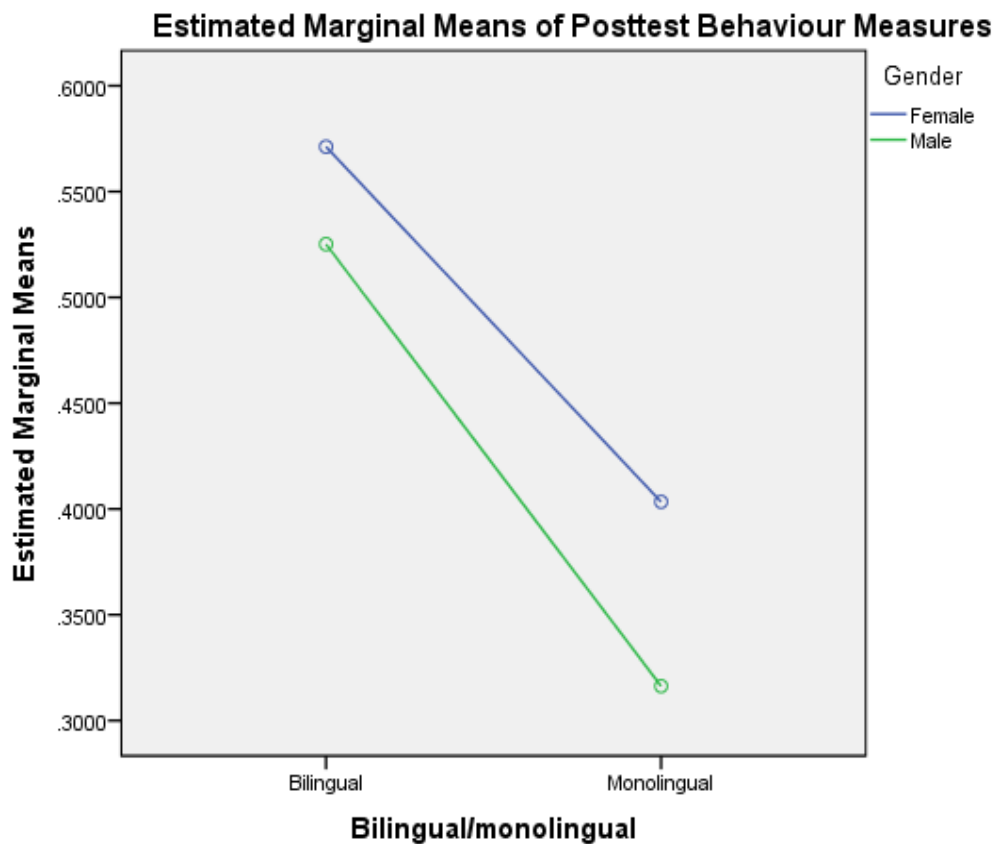
Table 10.3

Interaction Effect and Main Effect of Behaviour between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest Behaviour Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	134.304 ^a	4	33.576	43.280	.000	.183
Intercept	35.963	1	35.963	46.356	.000	.056
PretestBehaviourRaschMeasureon PosttestScale	106.191	1	106.191	136.881	.000	.150
Bilingualmonolingual	6.670	1	6.670	8.598	.003	.011
Gender	.818	1	.818	1.054	.305	.001
Bilingualmonolingual * Gender	.081	1	.081	.104	.747	.000
Error	601.235	775	.776			
Total	899.066	780				
Corrected Total	735.539	779				

a. R Squared = .183 (Adjusted R Squared = .178)

The main effect for the language program was statistically significant with $F(1,775) = 8,598, p = 0.003$; while that for gender was not, with $F(1,775) = 1,054, p = 0.305$ (see Figures 10.3 and Figure 10.4).

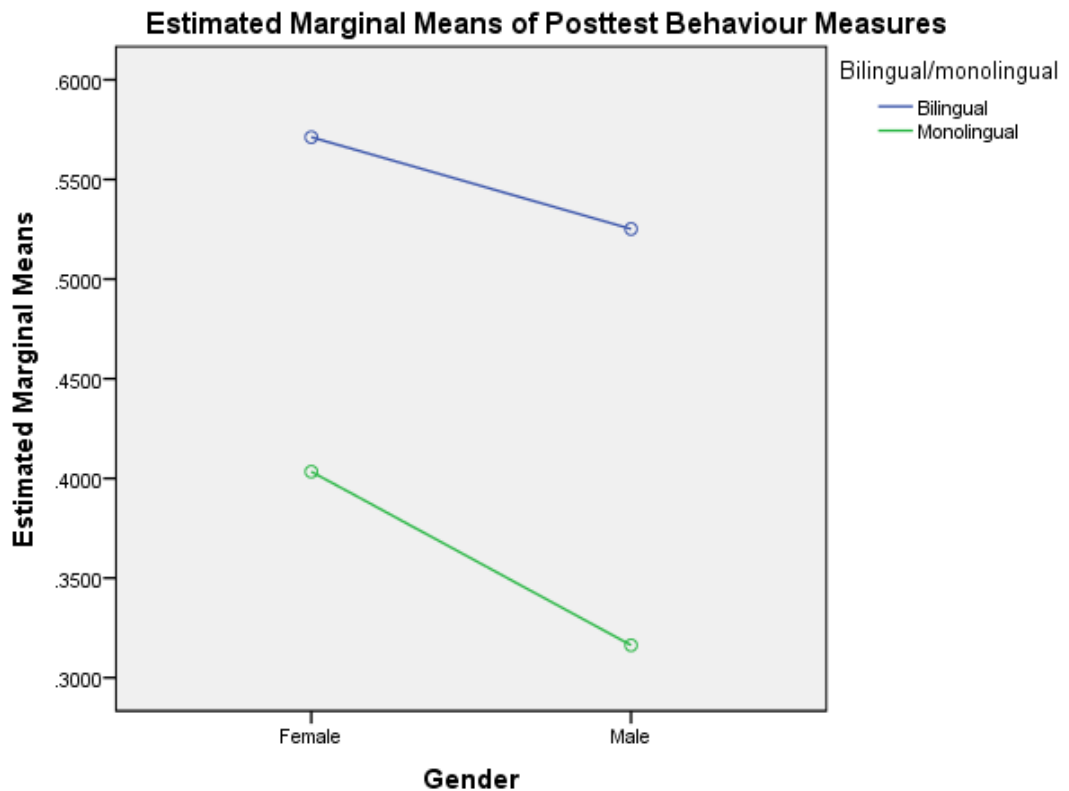


Covariates appearing in the model are evaluated at the following values: Pretest Behaviour Rasch Measure on Posttest Scale = .471579

Figure 10.3 Graph of Behaviour by Group (Bilingual/monolingual) with Pretest as Covariate

These ANCOVA results are different from the Split-Plot ANOVA results because the common covariate pretest measures were used to control for individual differences in monolingual and bilingual measures in one case, and in gender differences in the second case.

These results suggest that female students responded better than male students; and female and male students in the bilingual program responded better than female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest Behaviour Rasch Measure on Posttest Scale = .471579

Figure 10.4 Graph of Behaviour by Gender with Pretest as Covariate

Summary

Bilinguals (N = 394) and monolinguals (N = 386) students were given an English Behaviour Questionnaire pretest and posttest within a two month period (equivalent to 16 meetings), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA using the SPSS computer program. There are two main sets of findings: (1) with regards to the Mixed Design, and (2) with regards to the Two-Way ANCOVA.

The main findings for Behaviour with regards to Learning English, in respect of the Mixed Design ANOVA, are:

1. There is no significant interaction effect between bilinguals and monolinguals with $F(1, 775) = .104, p > .005 (.747)$, and a small effect size (partial eta squared = 0.000).
2. Bilinguals responded better than monolinguals.
3. Females responded better than males.
4. The present study suggested that the behaviour of girls with regard to learning English was important and that it, at least partially, may explain the well-known finding that English achievement of girls is better than that of boys.

The main findings for Behaviour with regards to Learning English, in respect of the Two-Way ANCOVA, are:

1. There was a main effect between bilinguals and monolinguals, with $F(1, 775) = 8,598, p = 0.003$, and a small effect size (η_p^2) = 0.011 (favouring bilinguals).
2. The main effect for gender was not statistically significant with $F(1, 775) = 1,054, p = 0.305$, and a small effect size (η_p^2) = 0.001.

The findings showed that bilingual students scored higher on the Behaviour measure than monolingual students. Female students from bilingual and monolingual schools scored higher on the Behaviour measure than male students from bilingual and monolingual schools.

The next chapter analyses the qualitative data on the students' written comments.

CHAPTER ELEVEN

DATA ANALYSIS (PART 7)

STUDENTS' WRITTEN COMMENTS

This chapter presents a qualitative analysis of the written comments on students' experiences with regards to learning English as a foreign language, using the Miles and Huberman (1994) method.

These data enrich the previous findings reported in Chapters Five, Six, Seven, Eight, Nine, and Ten. The present chapter analyses the opinions and thoughts of the students regarding their experiences in learning English prior to their participation in this study. Ten themes (or main issues) were identified: (1) Tasks on Learning English as a foreign language, (2) Student-Student Relationships, (3) Student-Teacher Relationships, (4) Personal Views on the Benefits of Learning English, (5) Common Views on the Benefits of Learning English, (6) Confidence and Achievement in Learning English, (7) Learning English through Media, (8) Family Support in Learning English, (9) Obstacles in Learning English, and (10) Other Views. These data give insights into how students feel with regard to their learning of English as a foreign language. Did they like learning English? What are the benefits of learning English? Did family, peers and teachers help them in learning English? Were there any obstacles in learning English? Did they have any support in learning English?

The questionnaire data that were collected between January and April 2011 included an additional written question as part of the posttest questionnaire. These written comments were provided in response to one written question on their questionnaire posttest, "***Is there anything else you would like to add about your experiences in learning English?***" Of the 780 students (bilinguals, N=394, bilinguals, N=386) who completed the posttest, 702 students (90%) answered this additional written question.

From the 702 students who answered the additional written question, 1842 written comments were provided: 1110 comments were from bilinguals and 732 were

from monolinguals. The comments, originally written in Bahasa Indonesia, were translated into English, back-translated to ensure the accuracy, and carefully grouped into ten areas and tally-marked manually. Tally marks are strokes written to record frequency of data, e. g. (////) and recorded as number of occurrences (see Appendix Q for the complete number of occurrences of written comments). A number of the students' best quotations that show comprehensive or unique information are provided here. By comprehensive information, it is meant that the comments were provided with details; and by unique, it is meant that the comments portrayed different senses and nuances that were not commonly depicted by other students. The students' comments were coded using their student numbers, which were allocated anonymously, and they are quoted verbatim.

It is worth noting that there are no gender-specific pronouns in Bahasa Indonesia. Both 'she' and 'he' in English are both referred to as '*dia*' in Bahasa Indonesia. On the written quotations, some students mentioned '*dia*'. Whenever the gender was unknown to the researcher, based on what the students provided, the pronoun 'she' or 'he' from the comments was translated to both 'she/he', for possessive pronouns to 'her/his'; and to objective pronouns to 'her/him'.

The analysis of the data discovered that both bilingually and Monolinguals found learning English as a foreign language is fun, yet challenging. However, they were motivated to learn it in order to help them for their future studies, for their future jobs and even to help them to go abroad and to study abroad. They thought that it is important to learn it, so they tried ways to make it easier to learn, despite obstacles along the way. Teachers helped in their journey of learning English.

Theme 1: Tasks on Learning English as a foreign language

There were four tasks on learning English as a foreign language identified: (1) Tasks for Listening, (2) Tasks for Speaking, (3) Tasks for Reading, and (4) Tasks for Writing.

Tasks for Listening

Some of comments from both groups stated that they liked listening to English songs (bilinguals, N=25; monolinguals, N=19). Other students said that they asked others to

speak slowly or repeat words in English (bilinguals, N=3); paid attention to someone speaking English (bilinguals, N=1; monolinguals, N=1); and that listening to English helped her/him to write (bilinguals, N=1).

Tasks for Speaking

More bilinguals than monolinguals (bilinguals, N=21; monolinguals, N=10) liked speaking in English. Eighteen comments showed that the students liked to practise speaking with other students, sang songs and believed that it was important to converse in English fluently.

Tasks for Reading

Some comments indicated that the students like learning English through reading (bilinguals, N=16; monolinguals, N=8). Only a small number of bilinguals, stated that they guessed the meaning of English words in the texts, and that they liked reading books/e-books in English.

Tasks for Writing

There were fewer comments on English Writing and only ten comments to cover five statements, affirming that students liked writing, wrote feelings in a diary, liked to create a new sentence with a new vocabulary, used a dictionary for new words, and understood written text better than spoken utterances.

The followings are students' quotations relating to Tasks on Learning English as a Foreign Language.

Students' quotations:

In our VII-1 class (seven-one), we were applying an English Day program which was held on Wednesdays. Sometimes, we forgot to speak in English, but we reminded each other. Our class was therefore called COSEVEN (Community of Seven-One). (Student # 1353, bilingual)

We have even got an English teacher from an English speaking country. The teacher taught me by using English language. I felt that I could understand him a little bit. From that time on I am motivated in learning English, so I will be able to communicate to people from other countries in English when I am overseas. (Student # 1022, monolingual)

So much so I wanted to be able to learn English, I sacrifice my afternoon time to take an after-school English tuition. And I do not mind to memorise English words which I do not know the meanings of. I also like to listen to songs in English. But I dislike getting questions in English, since I do not know their meanings. (Student # 1063, monolingual)

When I was having some walks, I would look at some items around me. I would try to name them in English. Should I do not know their names in English, I would seek help from my teacher or my friends. (Student #1384, monolingual)

I have been in a boarding school. Every day after *Shubuh* [an hour before sunrise] prayer, we were taught numbers of English words. I, then, used those words to communicate with my friends. I loved speaking English with my friends. I like it (Student # 1386, monolingual).

I was so happy when I communicated with some Dutch men in English (Student # 1109, bilingual).

Theme 2: Student-Student Relationships

Twenty-nine comments revealed that bilinguals learnt more when studying English in groups, and they asked friends to help them in learning.

Students' quotations:

I like English because I like to listen to it and learn it. Sometimes my friends and I read English books. We understand English and are able to write it and sing English songs. (Student # 1033, monolingual).

My experience in learning English is that I really want to speak and write well in English. But that is not happening yet. When I was in my primary school, my English lessons were not quite deep yet, so I asked a friend of mine to help me in English, because her/his English was quite good due to her/his school at one of favourite schools. (Student # 1021, bilingual).

I sometimes find English both confusing and fun. I know that it is impossible to understand it quickly, but it needs a process to learn it. I often seek help from my friends. (Student # 1048, bilingual).

I played games with my teacher and my friends. (Student # 1021, bilingual).

I have got a nice experience in learning English: I have performed an English dialogue with my friend in front of the class. (Student # 1116, bilingual).

I quite like English. I like it when I performed English dialogue with my friends in front of the class, and, I disliked it if I was to memorise in English. (Student # 1142, bilingual).

I have spoken in English with my friend. (Student # 1185, bilingual).

For me, learning English is lots of fun because my friends and I usually practice speaking in English. But sometimes, it turns out boring as well. (Student # 1188, bilingual).

My friends and I can speak English because we often practise speaking in English. My hope to study English is that in order it helps me to go overseas and study there. I also want to make my parents proud and happy. (Student # 1196, bilingual).

I got confused the first time I learn English. After some months I start to like it. And now I practise English quite regularly with my friends from the same year or from above. (Student # 1205, bilingual).

To sum up, both bilinguals and monolinguals who commented on this theme stated that they learnt English more when they studied it in groups. In addition to that, most of students commented regarding this theme asked friends to help them in learning English.

Theme 3: Student-Teacher Relationships

Many students from both groups (N=116) stated that they liked the way their teacher taught them English in that they learnt a great deal from their English teachers and they would seek help if they encountered problems. However, a small number (N=16) from both sides found the way their teachers taught English was not favourable to them. They also believed that the way teachers taught influenced their love or dislike of English.

Students' quotations:

Like the way English is taught

In my opinion, the way my teacher teaches me influences the way I learn English. Should the teacher be good, kind, and does not like to easily get grumpy, I like her. Frankly to say, should the teacher be mean and unkind; anything she/he has taught me will not reach my mind. (Student # 1134, bilingual).

My experience in learning English is enjoyable, because I like the way my English teacher teaches English. She makes me understand. (Student # 1364, monolingual).

I like my English teacher best. She is kind and not easy to get angry. (Student #1371, monolingual).

Dislike the way English is taught

I am a little bit confused with my English teacher because he does not understand English well. The way he writes, reads and explains English makes me dislike to be taught by him. When we were learning English, the teacher explains it in local language. We got mixed up. Therefore, I learn again at home or at my English tuition centre so my English knowledge and skills keep increasing. (Student # 1026, monolingual).

I like English but I do not like the way my teacher teaches it. Because of his teaching at my school, I do not like English anymore. (Student # 1028, monolingual).

As a matter of fact, I prefer overseas teacher to teach us English. It seems easier for me to understand, and it would be easier if the teacher can also speak in Bahasa Indonesia. (Student # 1243, monolingual).

Yes. Sometimes my teacher does not answer some questions that I asked. (Student #1363, monolingual).

As a summary, most bilinguals and monolinguals who commented on this theme liked the way their teachers taught English and only some did not like their teachers' methods. Some of them learnt a great deal from their English teachers and were pleased to state this but some other students believed that the way a teacher teaches influences whether students like or dislike learning English.

Theme 4: Personal Views on the Benefits of Learning English

Students commented that they liked learning English (bilinguals, N=96, monolinguals, N=70); that English was fun (bilinguals, N=76, monolinguals, N=50); that it was challenging (bilinguals, N=19, monolinguals, N=9); that learning English is important (bilinguals, N=9, monolinguals, N=12); that it was easy (bilinguals, N=15, monolinguals, N=5); that English was a unique language (which sounds nice but is is

difficult to pronounce) (bilinguals, N=7, monolinguals, N=9); that they were aware that a number of good texts were written in English (bilinguals, N=1); that they preferred to learn English with simple sentences (bilinguals, N=1, monolinguals, N=2); that English was their favourite subject (bilinguals, N=1); that learning English needed time (bilinguals, N=2, monolingual, N=1); that other languages helped them understand English (bilingual, N=1, monolingual, N=1); that they liked storytelling (bilinguals, N=2); that they liked creating stories in English (bilingual, N=1); that they liked to translate English into Bahasa Indonesia and vice versa (monolinguals, N=3); and that an English dictionary helped them (bilinguals, N=7, monolinguals, N=3) (see Table 11.1).

Table 11.1
Students' Personal Views on the Benefits of Learning English as a Foreign Language

Personal Views on Benefit of Learning English as a Foreign Language	Bilingual	Monolingual	Total
I like learning English	96	70	166
I like English because it is enjoyable/fun	74	50	124
I like English because it is challenging	19	9	28
Learning English is important	9	12	21
Learning English is easy	15	5	20
I like English because it is unique (it sounds nice but is difficult to pronounce)	7	9	16
Using a dictionary helps me	7	3	10
I am happy with my vocabulary	1	3	4
I want to learn English with simple sentences	1	2	3
Learning English needs time	2	1	3
I like to translate	0	3	3
English is enjoyable if studied correctly	2	0	2
Other languages help me understand English	1	1	2
I like storytelling	2	0	2
I like English because a number of good books are in English	1	0	1
English is my favourite subject	1	0	1
I like creating stories in English	1	0	1

Source: Prepared by the author from the students' written answers

Students' quotations:

I used to dislike English lessons. In every English session at my primary school, my English teacher was always angry with me. Now, I am at this Middle School with a bilingual program, so I start loving English. (Student #1274, bilingual)

First, I like English if it is used in daily life. Second, I like English if there is somebody who can translate it into Bahasa Indonesia well. (Student # 1394, bilingual).

English is to increase my knowledge, to be able to read comics and novels, to understand songs in English, and to help me for the future. (Student # 1006, monolingual).

English teaches me to start a new life. It could help me. English is like a smile in my heart that gives blessing in my heart. (Student # 1141, monolingual).

I wish all Indonesian people to be able to speak English. Amen. (Student # 1183, monolingual).

At first learning English, it was boring enough, but gradually it turned out quite fun and enjoyable, because I started to understand it. Then?... I like it. (Student # 1303, monolingual).

My experience in learning English is that I am trying to enjoy the lessons even though I do not like English. I am trying to understand words in English. (Student # 1375, monolingual).

I often converse in English alone at home. I often speak with my younger siblings, for example: 'eat', 'drink', in English. (Student # 1360, bilingual).

When I was at kindergarten, I used to be afraid that I could not read and understand English. But as time goes, my knowledge of English builds up. At present I often use English in my daily life. (Student # 1363, bilingual).

English is fun but at the same time it is annoying. However, thinking deeply, I prefer English to my local language (Acehnese). It is easier to learn English than Acehnese. (Student # 1072, monolingual).

Yes, of course. When I was still poor in English, one overseas man asked me, "where do you live"? I was young at that time and could only answer, "I don't know", and shook hands with him confidently. (Student # 1296, monolingual).

Summarising this theme, the majority of both bilinguals and monolinguals stated that they liked learning English because it was fun and enjoyable, easy, important, unique (that is, pleasing on listen but difficult to pronounce), yet challenging. Some students also stated that dictionaries helped them. They were happy with their

vocabulary, they wanted to learn English with simple sentences, and they believed that was enjoyable, if studied correctly, but that needs time to learn.

Theme 5: Common Views on the Benefits of Learning English

Many students (bilinguals, N=77; monolinguals, N=66) stated that the benefit of learning English was that it could help them to go abroad. One hundred and thirty-five comments said that the students learnt English to help them in higher study (bilinguals, N=76; monolinguals, N=59); while 87 students stated that they needed to learn English because it was an international language (bilinguals, N=46; monolinguals, N=41), and 75 commented that English was a means to pursue dream jobs (bilinguals, N=46; monolinguals, N=27). Twenty-five comments showed that English helped them in their daily activities (bilinguals, N=16; monolinguals, N=9) and 25 comments revealed that English could help them study abroad (bilinguals, N=17; monolinguals, N=6). The rest stated that English was regarded as an additional language (N=10) (bilinguals, N=3; monolinguals, N=7); that it was used in the classroom (bilinguals, N=6); that it could help them to get enrolled in favourite schools (N=3) (bilinguals, N=2; monolinguals, N=1); and that it could help them to get a scholarship for higher studies (bilinguals, N=1) (see Figure 11.1).

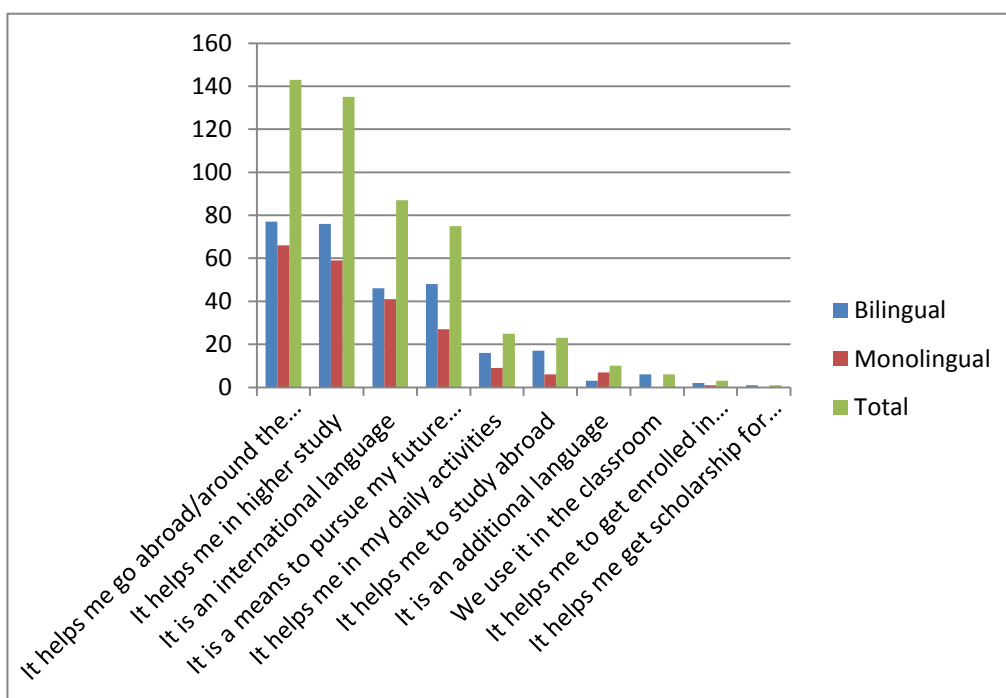


Figure 11.1 Students' Common Views on the Benefits of Learning English

Students' quotations:

Many students were very positive about learning English:

I learn English in order that I become a successful person and enable me to understand plenty of problems in English, and also to be able to work with important people overseas when I become a successful person. (Student # 1071, bilingual).

I like learning English because it could help me to study in Sydney. I really want to study medical sciences there. I have been learning English in many ways, such as listening to English songs; films; and story-books. I also learn English with the help of my Mum, because she is an English and German teacher. My favourite song is 'Hush Hush'. (Student # 1168, bilingual).

I like English, because with that I am free to go to any country with different languages (by using English), also I do not need to bother with the languages used in those countries. (Student # 1191, bilingual).

My experience in learning English is that I know English now. English is one of the international languages. If we go overseas, if we do not know English, we won't understand anything, and we cannot do anything. But, if we know English, we can easily understand what the people are talking about, and we can easily acquire whatever we need overseas. (Student # 1203, bilingual).

When I was in Singapore on holiday, I got lost. I wanted to ask the way to get back to my hotel, but they did not understand me because they used English. Fortunately, I know English a little bit. So by learning English we can understand it. (Student # 1207, bilingual).

English is so enjoyable. Especially in the globalisation era, it is a must for us to speak English fluently. Because nowadays, both Mathematics and English are the most important subjects, so automatically we have to be fluent in English. (Student # 1335, bilingual).

In my opinion, English is the coolest subject to learn because if we are able to master English, it will be needed and we can pursue our study overseas. (Student # 1393, bilingual).

I like English, because when I encounter some overseas people having a holiday in my city, I can speak to them. (Student # 1047, monolingual).

I really like English because I want to be a medical doctor working overseas. (Student # 1065, monolingual).

I like learning English, because if I can speak English, I can speak with foreigners when I am older, and also can travel around the world. If we can

Speak English it could help us get a good job, and we can even work overseas. (Student # 1279, monolingual).

In my experience, I am happy because with English I can speak with foreigners or can understand what they are talking about. (Student # 1345, monolingual).

One student was initially negative about English, but then changed

Actually I did not like English, but interestingly, now I start to like it, because if I can speak English I can pursue my higher study overseas. (Student # 1377, bilingual).

To sum up, most of the students learnt English because they believed that it would help them to go abroad, or around the world, and help them in higher study later. The students believed that English was worth studying because it is an international language and it is a means to pursue a dream job. Some other students liked English as an additional language to help them in their daily activities.

Theme 6: Confidence and Achievement in Learning English

Many students (bilinguals, N=151, and monolinguals, N=87) stated that they were motivated to learn English. Some students (bilinguals, N=18, and monolinguals, N=7) took after-school English tuition to improve their English. Other students said that they had won an English competition (bilinguals, N=11; monolinguals, N=7), were happy with their vocabulary (bilinguals, N=1; monolinguals, N=3), liked learning English from a very young age (bilinguals, N=11), and felt lucky to understand English (bilinguals, N=7).

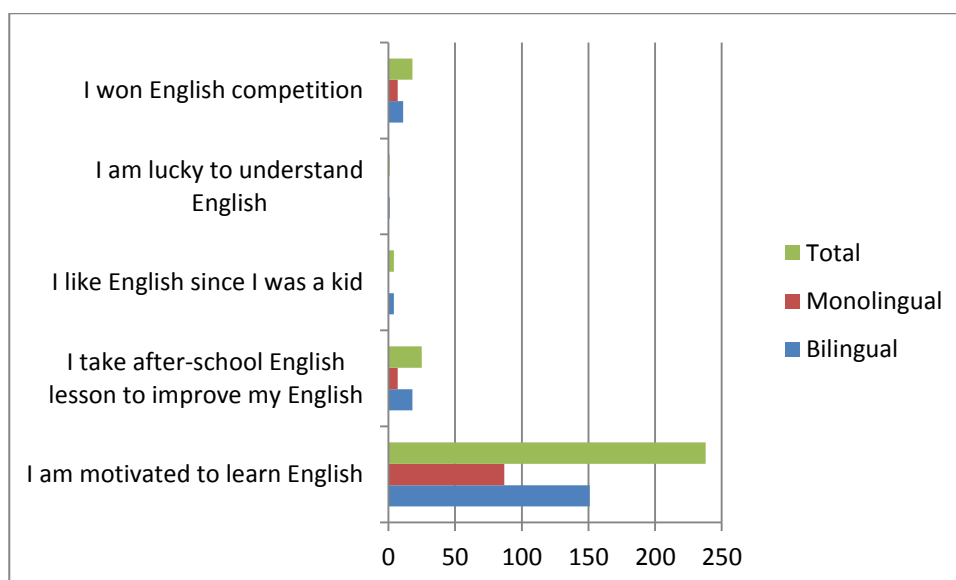


Figure 11.2 Students' Confidence and Achievement

Students' quotations:

I was so happy when taking part in a story-telling competition. I did not win anything at that time but my brother did. I was happy enough because my English teacher said that my pronunciation was good. I guess I have to study English harder. (Student # 1111, bilingual).

First time studying English I could only say, 'yes/no'. But from that day on I keep learning English. It turns out that English is fun. Moreover English is a famous language and is used daily. (Student # 1299, bilingual).

Yes. It took me quite some time until I understand English. Actually it does not work if we learn English only at school; therefore, I study it again at home with the help of my older sister and my parents. Gradually I understand it. (Student # 1354, bilingual).

I prefer when it has songs as well. I am happy with English and I am quite fluent in it. I am good at it now because I like to watch Japanese films with English subtitles. I also like western films. (Student # 1389, bilingual).

I like English so much; especially when I lead my friends reading some texts on the whiteboard, WOW... It feels like I am an honour person. I would like to be a professional English teacher when I grow up. (Student # 1057, monolingual).

I like English very much, because my first and prime wish is to be able to speak and write in English well, clearly and correctly. (Student # 1061, monolingual).

I have experience in learning English. At this age, I am able to speak English and use it within my school. However when I used it in my classroom, some friends said, “you show off with your English”. Probably they said that because they do not understand English. (Student # 1381, bilingual).

To summarise, the majority of respondents from both bilingually-taught and monolingually-taught schools responded that they were motivated to learn English. They took after-school lessons or tuition and some were happy because they had won some English competitions.

Theme 7: Learning English Through the Media

More bilinguals (N=27), commented that they liked playing games, singing songs, watching TV programs, and watching English films, either to understand, maintain or improve their English. The rest (N=9) stated that they were aware that number of media, software, and computer games used English. Therefore, they wanted to know English in order to use those media correctly. Of the comments, only one revealed that she or he preferred learning English through his or her school laboratory.

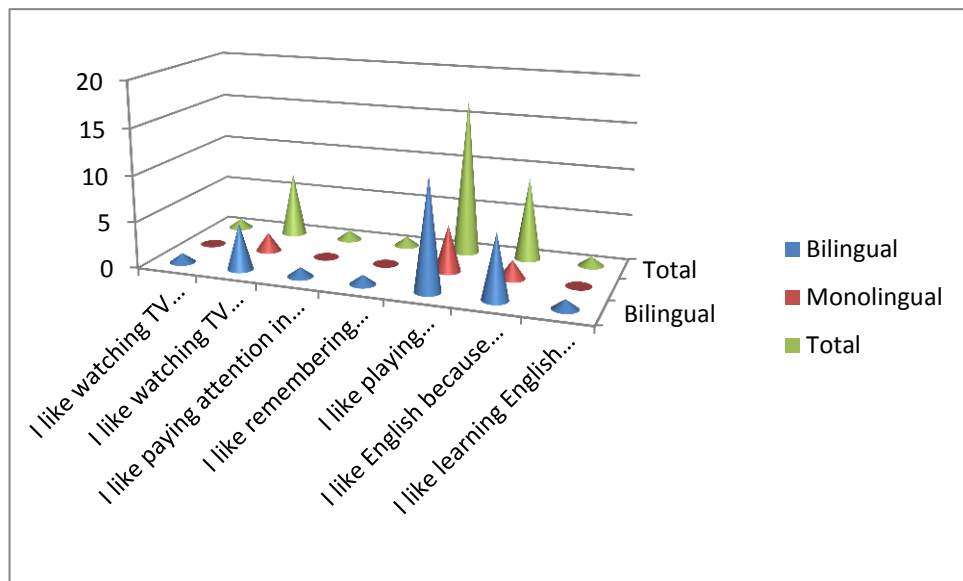


Figure 11.3 Learning English through the Media

Students' quotations:

Yes. Every Sunday and on public holidays I am learning through Australia Radio. The session has 104 lessons for 2 years. I learn English to ease my way to

go abroad and to study in the Netherlands or Australia. (Student # 1310, monolingual).

I like English because it is an international language and almost all games and software are in English. (Student # 1365, bilingual).

I like English for my future and in order to know the content of games at my house. (Student # 1378, monolingual).

I love English because it could help me to understand English in games as well as my daily life. (Student # 1382, monolingual).

I like English. I also like to listen to English songs and watch western movies. Learning English is a lot of fun. (Student # 1386, monolingual).

I like understanding English through English-based videos. (Student # 1031, monolingual).

I often watch western films with English subtitles when learning English. Should I find a new word, I would take a note, then refer to a dictionary. I would retell my English teacher the story on the following day. (Student # 1159, monolingual).

I often watch films and songs in English. It really helps me to increase my English vocabulary. (Student # 1175, monolingual).

I was once invited to talk with my overseas teacher at an English tuition. She/he also asked me to play games. She/he spoke so fast that I did not understand. (Student # 1183, monolingual).

To sum up, some students liked playing games in English, listening to English songs, and reading English story books in order to improve their English. They were also aware that numerous forms of media, software, games, computer programs, or videos use English.

Theme 8: Family Support in Learning English

There were 43 students who commented on family support (bilinguals, N=37; monolinguals, N=6). Their family, that is, fathers, mothers, uncles, aunts, and siblings encouraged them to learn English. Thirty-three bilinguals liked to practise English conversation with their older brothers or sisters. The students emphasised that they learnt English well in order to make their family proud. A few commented

(monolinguals, N=7) that one of the reasons they learnt English was because they wanted to teach English to their family members.

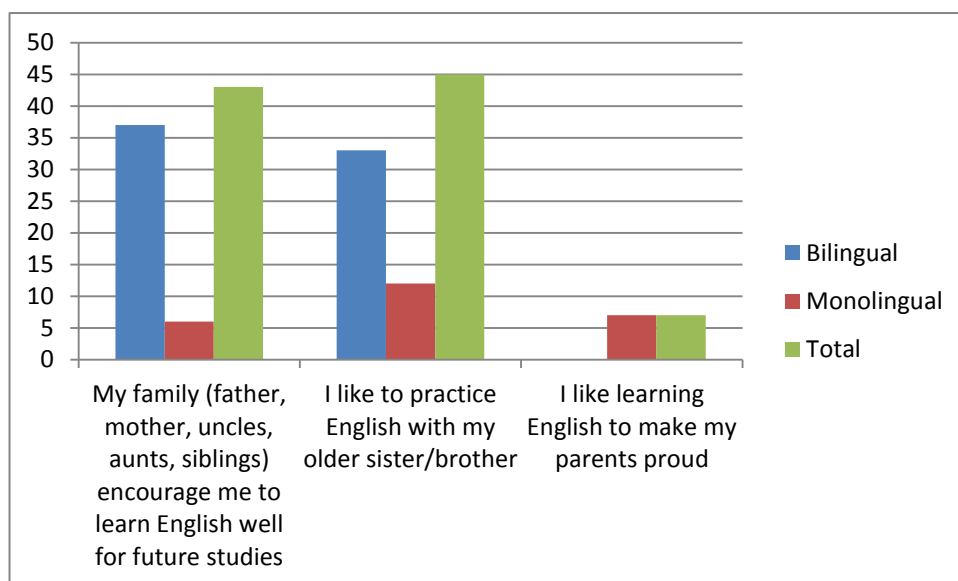


Figure 11.4 Family Support in Learning English

Students' quotations:

My parents often help me learning English, for example, my mum often ask me to communicate in English; my father often buys me English storybooks. A couple of my friends on Facebook are living overseas, and I love chatting with them in English. (Student # 1251, bilingual).

Yes. I did not like English. But, because my mother is an English teacher and I listen to her talking English with my bigger sister, then, I am interested in it. (Student # 1393, bilingual).

English can make us smart and successful, and can also make our parents proud. (Student # 1062, bilingual).

I like English very much because it could help me go overseas to be an English interpreter and can add some value to me. My parents will be so proud of me if I can understand English. (Student # 1349, bilingual).

I learn English with members of my family. It happens that my mum and my cousin know English a little bit. I like English because it is fun, and cool (Student # 1282, bilingual)

I love English very much. I have always got good scores. Thanks to my teacher and my elder sister who is studying to be an English teacher. I often ask her words/meanings that I do not know about. I also usually speak English with her. (Student # 1371, bilingual)

I like learning English with my teacher so much. I also learn it with my elder sister. (Student # 1102, monolingual)

It is easy to learn English as long as we learn it seriously. Also, we will be able to use it soon. English is the most interesting. (Student # 1113, monolingual)

When I was 4 years old, I got help from my mum, sister, and cousin (who was an English lecturer). I often seek help from her and she helps me much. (Student # 1151, monolingual)

I like English lesson because I was taught to speak English when I was 5 years old. Thanks to God, I love English so much now. It started from my mum. I got motivation from my family who are already fluent in English; therefore I have to be fluent in English, too. (Student # 1060, monolingual)

This theme summarises that some students received help from their parents and older siblings, as well as from other family members regarding the learning of English. They also liked to learn English in order to make their family proud of them.

Theme 9: Obstacles in Learning English

Some students claimed English was a difficult subject (bilinguals, N=31, Monolinguals, N=41). They said that vocabulary, meanings and grammar were not easy to understand (bilinguals, N=15, monolinguals, N=21). English was not easy to learn (bilinguals, N=6, monolinguals, N=22) and the four skills, that is listening, speaking, reading, and writing, were difficult to acquire (N=34 for both). Four students stated that they did not like English because English was not needed in Indonesia. Only eight students (bilinguals, N=3, monolinguals, N=5) wished to know English better; and seven students (bilinguals, N=5, monolinguals, N=2) were aware that they were not good at English.

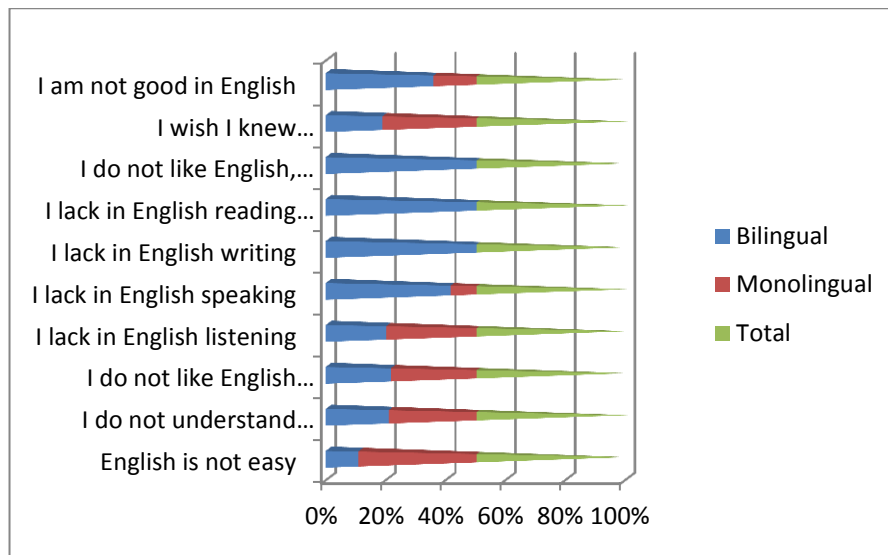


Figure 11.5 Obstacles in Learning English

Students’ quotations:

I like English, but I dislike it a bit, because when writing it, the words written are different to the way they are pronounced. I find it difficult when we do dictations with my English teachers. (Student # 1180, bilingual).

I love learning English because English can help me do higher study. But sometimes it is difficult. The most difficult ones are on ‘pronouns’, ‘possessive adjectives’, and ‘possessive pronouns’. (Student # 1313, bilingual).

My English is not fluent because I do not like English. It is so difficult for me, so I like sports instead. However English is also important to help me for future studies and when I am overseas. (Student # 1009, monolingual).

No. I do not like English much. (Student # 1012, monolingual).

In my opinion, learning English is fun, however, I do not know why I do not like it. (Student # 1284, bilingual).

There are plenty of grammar/rules in English. They make me so confused. (Student # 1292, monolingual).

Actually I like English, but sometimes I do not know why I do not understand what people are saying in English. I think I won’t be able to speak English even though I have been crossing the seas. Thank you. (Student # 1043, bilingual).

I do not like English because there are plenty of English words that I do not know. That sometimes make me find English annoying. Sometimes I do not know what people are talking about English are. (Student # 1148, bilingual).

I do not like English so much because it is difficult to understand. Besides, we do not communicate in English in our country. (Student # 1139, bilingual).

As a summary, English is viewed as a difficult additional language for some students, in terms of vocabulary, grammar, and meanings. A few students were aware that they lacked ability in listening, reading, and writing and others wished that they knew English better.

Theme 10: Other Views

Five students found that the tests conducted by the researcher were beneficial and wrote that the tests gave them good insights into English lessons. They were considered to be the first tests that were well prepared and presented.

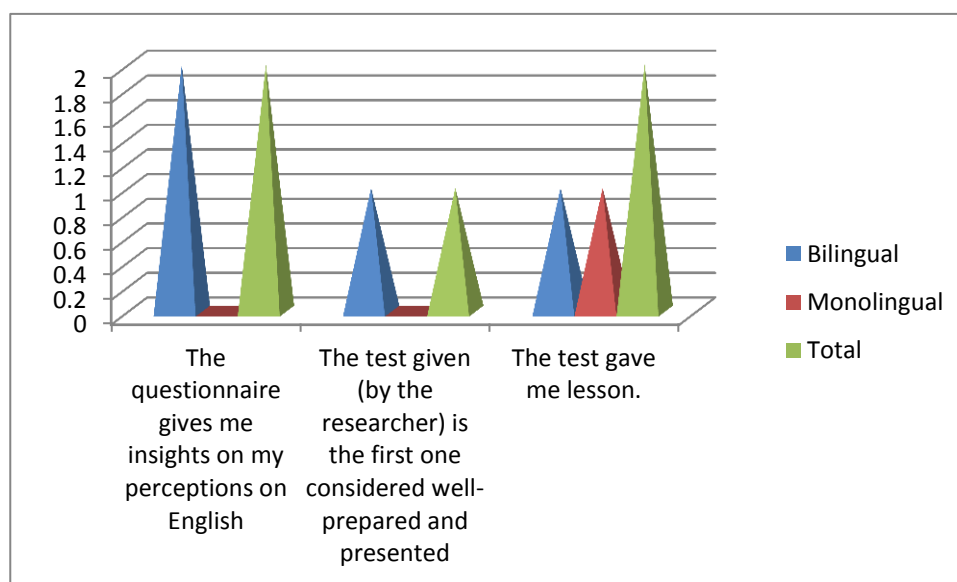


Figure 11.6 Other Views

Students' quotations:

I have been experiencing good English experiences when it was taught by Mr. R in this school. And this is the very first well-presented Reading Comprehension test that I have ever had. (Student # 1061, bilingual).

Thank you. With this questionnaire, I can tell my feelings, and through this questionnaire I know more than before. I like answering this questionnaire. (Student # 1077, bilingual).

Summary

Using the Miles and Huberman (1994) method, ten themes were created from 1110 comments from bilingual students and 732 comments from monolingual students. Both bilinguals and monolinguals shared similar opinions about the ten themes. The students' written comments generally expressed their positive attitudes with regard to learning English as a foreign language. Most of the students thought that learning English was important for their present and future needs. Even though they faced some obstacles in learning English, they kept themselves motivated, and stated that they wanted to continue to learn English as well as Indonesian.

The results of the qualitative analysis added more information to the results of the quantitative analysis. Both analyses provided new knowledge on Acehnese students' abilities in English reading comprehension, English writing, and students' attitude and behaviour, as well as on the students' opinions and thoughts regarding their experiences in learning English as a foreign language.

The next chapter answers the research questions, and then discusses the results and the implications of the present research study.

CHAPTER TWELVE

DISCUSSION OF RESULTS AND IMPLICATIONS

This concluding chapter presents an overview of the study conducted, gives answers to the research questions, provides a summary of the findings, as well as presents discussion about bilingual education. It also provides some limitations and implications of the study and gives some recommendations for further research. The present study was aimed at investigating whether students who were learning English bilingually through the combined medium of English language and Bahasa Indonesia were better in some English abilities (English Reading Comprehension and English Writing) compared to students who were learning English with the medium of Bahasa Indonesia only. It also sought a comparison between students learning English bilingually and students learning English monolingually in their attitude and behaviour with regard to learning English as a foreign language.

I live in Banda Aceh and, from anecdotal evidence, I know that the perceptions and beliefs of the Acehnese community, consisting of parents, students, and teachers are that bilinguals at state schools perform significantly better than their counterparts. These perceptions and beliefs seemed to occur after some academic results at certain schools following the implementation of the new 2004 English curriculum that required every state school, from elementary level to secondary level, to provide some classes with bilingual programs.

In Aceh Province, the implementation of the new English curriculum was supported by a number of international workers who were welcomed to help reconstruct and rebuild Aceh after the 2004 Tsunami catastrophe. During that time some international organisations established bilingual boarding schools, where only English and Turkish were used. These schools soon gained in popularity.

Popularity was not only enjoyed by the Turkish boarding schools, but also state schools with bilingual programs. Large numbers of parents wanted to enrol their

children at the state schools with bilingual programs. Teachers also encouraged parents and student candidates to enrol at bilingual schools.

However, there was no known research evidence as far as could be ascertained, in Indonesia that bilinguals outperformed monolinguals. There was not a single research study reported and conducted in Aceh Province, or in any other provinces of Indonesia, that investigated this issue. Therefore, the present study was conducted to investigate the answer as to whether or not state school bilinguals did perform better than state school monolinguals in Aceh Province.

Summary of the Present Study

The data collection for the present study was carried out in Banda Aceh, from August 2010 until April 2011. The pilot study was carried out in October and November 2010, and the main data collection was administered from January 2011 until April 2011. Thirteen state Middle Schools took part in the study, with 780 female and male student respondents, consisting of 394 students from bilingually-taught schools (N=5) and 386 students from monolingually-taught schools (N=8). The study involved control and experimental groups with pretests and posttests using three measures: (1) English Reading Comprehension; (2) English Text Writing; and (3) an Attitude and Behaviour Questionnaire. All schools share similarity in terms of English curricula, English syllabi, English lesson content, English lesson times, English lesson duration, homework, and text books, which then were taken into account for recruitment of student-respondents of the two types.

The 780 students sat for three tests: English Reading Comprehension, English Writing; and Behaviour Questionnaire, for both pretests and posttests. The data analysis for three variables was conducted with the RUMM2030 computer program (Andrich, et al., 2010), to create linear scales, and a Mixed Design ANOVA and Two-Way ANCOVA (SPSS) were conducted to test for differences between the control and experimental groups.

This dissertation consists of twelve chapters. *Chapter One* started by introducing readers to the Indonesian educational system and to bilingual and monolingual education in Aceh province, Indonesia. It also presented the rationale and

background behind the study, as well as its aims. *Chapter Two* briefly revisited the nature of bilingualism, the nature of bilingual education, including bilingual education in Indonesia, and summarised some of the more relevant and recent research (2003-2013) on these topics from countries across the world. The conceptual framework of bilingual and monolingual teaching, measurement, and the variables of the study were presented in *Chapter Three*. The chapter informed readers about the conceptual framework behind bilingual and monolingual teaching in Aceh province, Indonesia, and explained why bilingual education was expected to produce superior achievements over monolingual education after a two month experiment. It also presented the three dependent variables of the study: English Reading Comprehension, English Writing, and Behaviour with regard to learning English, and explained the conceptual structure of each variable. Then there also was an explanation of the problems of True Score Theory measurement and why it was not used in the present study, but a better alternative – Rasch measurement – was used instead. *Chapter Four* explained the design of the study, covering the research strategy and design; mixed-method design; samples; pilot study; study and ethics approvals; control of extraneous variables in the quasi-experiments, and test data collection, data entry, and analysis of the quantitative and qualitative data.

The findings of the study were presented in Chapters Five to Eleven. *Chapter Five* showed the findings of the Rasch analysis of the Reading Comprehension. *Chapter Six* reported the experimental results on Reading Comprehension. The results of the Rasch analysis of English Writing were presented in *Chapter Seven*. The findings of the experimental results on English Writing were shown in *Chapter Eight*. *Chapter Nine* reported the findings of Rasch Measurement of Behaviour. The results of the experimental findings of the Behaviour were showed in *Chapter Ten*. *Chapter Eleven* reported students' written comments with regard to learning English as a foreign language. *Chapter Twelve*, the concluding chapter, presented a discussion of the results and the implications of the present study.

Answering the Research Questions

The purpose of the present study was three-fold. The first was to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English

text writing, English reading comprehension, and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual/monolingual). The second was to investigate the behaviour of students who have been taught English bilingually and of those who have been taught English monolingually, in term of their ability in English. The third is to investigate the students' perceptions, through their written comments, in relation to their learning English as a foreign language.

The study's purposes were presented in four research questions, which guided the investigation and informed the methods used, as well as the analysis of the quantitative and qualitative data.

Research Question 1 *Do first year middle state-school bilinguals in Banda Aceh achieve better in English Reading comprehension and English text writing than those of monolinguals?*

Based on the Rasch-created linear scale for Reading Comprehension (see Chapter Five), the first year middle school bilinguals in Banda Aceh achieved better in English Reading Comprehension than monolinguals. There was a significant difference in output by bilinguals compared to monolinguals, with bilinguals achieving at a higher standard on most items. This is supported by a significant difference in overall Rasch measures for Reading Comprehension by type of teaching (either bilingually or monolingually), in favour of bilingual teaching.

The result for the Rasch analysis was supported by the experimental results using SPSS ANOVA and ANCOVA (summarised in Chapter Six). Bilinguals significantly outperformed monolinguals on English Reading Comprehension.

Results for English Writing were reported in Chapter Seven and Chapter Eight. Based on the Rasch-created linear scales, bilinguals had a statistically significantly better English Writing ability than monolinguals.

The results of the Rasch analysis (summarised in Chapter Seven) on English Writing was supported by the experimental results using SPSS ANOVA and ANCOVA

(Chapter Eight). Bilinguals performed significantly better than monolinguals on English Writing, with $F(1,775) = 124.69$, $p = 0,000$, with partial eta squared = 0.139.

Research conducted in countries other than Indonesia has supported the view that bilingual teaching produces superior results to monolingual teaching. A summary of research done over the last decade (2003-2013) supports this view.

Research Question 2 *Do first year middle state school bilingually-students at bilingual schools in Banda Aceh have better behaviours with regard to learning English than those at monolingual schools?*

Based on the results of the Rasch-created linear scale for Behaviour (Chapter Nine), first year middle school bilinguals in Banda Aceh had better behaviours with regard to learning English than monolinguals. This is supported by the difference in overall Rasch measures for Behaviour by type of teaching, which was statistically significant in favour of bilingual teaching.

The results of the experimental study were in agreement with this. There was a main effect between bilinguals and monolinguals, where bilinguals had better behaviours with regard to learning English as a foreign language.

Research Question 3 *What are the attitudes and behaviours of first-year middle state school bilinguals with regard to bilingual and monolingual education in Banda Aceh in term of learning English?*

A Rasch-created linear scale of Attitude and Behaviour with regard to learning English as a foreign language provided a list of behaviours ranging from the easiest for the students to perform, to the most difficult. They are listed here from easy to hard:

1. I actually like the way my teacher teaches English writing. (*easiest*)
2. I actually pay attention to someone speaking English.
3. I actually read words in English carefully.
4. I wish to say new English words several times.
5. I actually can understand English better when I do activities with friends.
6. I actually do read English at home on my own.
7. I actually learn more when I study English in groups.
8. I actually like English because we use it in the classroom.
9. I actually practise English with other students.
10. I actually say new English words several times.
11. I actually start conversations in English with my friends. (*hardest*)

The ten statements from the Behaviour and one statement from the Attitude list have been categorised into three levels of difficulty, based on the students' answers. Statement numbers 1-4 are returned as easy, 5-8 as quite difficult, and 9-11 as most difficult. Statements numbers 1-4, such as statement number (1) *I actually like the way my teacher teaches English writing*, fell under the category of 'easy' due to the fact that they were answered positively by the majority of the students. As the present study did not investigate the reasons behind the students' answers, it was assumed that the reason for answering those statements positively was because the activities involved students' personal willingness, which seemed easier to perform than statements with activities that involved other people (numbers 5-8), such as statement numbers (5) *I actually understand English better when I do activities with friends*, and (7) *I actually learn more when I study English in groups*.

Statements numbers 9-11 were considered difficult. There was only a small number of students who answered those statements positively, which was assumed to be because the stated activities involved active engagement in English, such as statement numbers (9) *I actually practise English with other students*, and (11) *I actually start conversations in English with my friends*.

Research Question 4 *What are first-year middle state school students' perceptions with regard to their learning of English as a foreign language?*

Results from the qualitative data (reported in Chapter Eleven) revealed that the students had positive perceptions towards their learning of English as a foreign language. The students' perceptions were categorised into ten themes. The following are the themes with statements the students mostly wrote for each theme.

Theme 1: Tasks on Learning English as a foreign language

- I like listening to English songs
- I like speaking in English
- I like learning English through reading comprehension
- I like writing

Theme 2: Student-Student Relationship

- I learn more when I study English in groups

Theme 3: Student-Teacher Relationships

- I like the way my teacher teaches English

Theme 4: Personal Views of Learning English

- I like learning English

Theme 5: Common Views on the Benefits of Learning English

- I like English because it helps me go abroad/around the world

Theme 6: Students' Confidence and Achievement in Learning English

- I am motivated to learn English

Theme 7: Learning English through the Media

- I like playing games/singing songs/reading story-books to improve my English

Theme 8: Family Support in Learning English

- I like to practise English with my older sister/brother

Theme 9: Obstacles in Learning English

- I do not like English because it is not easy to learn

Theme 10: Other Views

- The questionnaire gives me insights into my perceptions of English

Amongst the ten themes, the most frequently mentioned comments related to ‘Common Views on the Benefits of Learning English’, with 505 written comments, followed by ‘Personal Views of Learning English’ with 408 written comments, and ‘Students’ Confidence and Achievement in Learning English’, with 286 written comments. The least frequently mentioned themes were ‘Learning English through the Media’, ‘Student-Teacher Relationships’ and, ‘Other Views’, with 37, 29, and 10 written comments respectively.

Generally, the students’ written comments showed that they were very positive about learning English in relation to the way they learnt and experienced English. They liked learning tasks regarding the four English skills (listening, speaking, reading comprehension, and writing). They mentioned that they had good relationships with their fellow students and teachers. Even though some students faced difficulties in learning English due to its complexity in vocabulary, grammar, and meanings, most students revealed that they liked learning English because it was challenging and enjoyable, as well as helping them to go abroad and helping them in higher education.

The students also stated that they were motivated to learn English. Some students took after-school English lessons to improve their English, as well as learning through the media. In addition to this, they received good support and attention from their family. Some students mentioned that the tests that the researcher gave them helped give them insights into their perceptions of English.

Summary of the Findings

The main findings are set out under: (1) English Reading Comprehension, (2) English Writing, and (3) Questionnaire.

A. English Reading Comprehension:

1. Both the bilinguals and monolinguals in Banda Aceh achieved better scores in English Reading Comprehension posttests than in the pretests.
2. Bilinguals performed better in the pretest than monolinguals.
3. Bilinguals performed better in the posttest than monolinguals.
4. Female bilinguals performed better than male bilinguals.
5. Female monolinguals performed better than male monolinguals.

6. Female bilinguals performed better than female monolinguals.
7. Male bilinguals performed better than male monolinguals.
8. Generally, female students outperformed male students.
9. Generally, bilinguals outperformed monolinguals.

B. English Writing:

Some caution should be exercised in interpreting the English Writing results since two months between pretest and posttest measures maybe too short to obtain strong improvements.

1. Bilinguals performed better in the English Writing pretest than monolinguals.
2. Bilinguals performed better in the English Writing posttest than monolinguals.
3. Female bilinguals performed better than male bilinguals in English Writing.
4. Female monolinguals performed better than male monolinguals in English Writing.
5. Female bilinguals performed better than female monolinguals on English Writing.
6. Male bilinguals performed better than male monolinguals in English Writing.
7. Generally, female students outperformed male students in English Writing.
8. The posttest results for English writing were measured to be lower, after two months, than the pretest result, and this is true for both females and males, and for both bilinguals and monolinguals.
9. In general, bilinguals outperformed monolinguals in English Writing.

C. Behaviour with regard to Learning English as a Foreign Language:

1. Bilinguals scored higher than monolinguals in the pretest Behaviour measure.
2. Bilinguals scored higher than monolinguals in the posttest Behaviour measure.
3. Female bilinguals scored higher than male bilinguals in the Behaviour measure.
4. Female monolinguals scored higher than male monolinguals in the Behaviour measure.
5. Female bilinguals scored higher than female monolinguals in the Behaviour measure.
6. Male bilinguals scored higher than male monolinguals in the Behaviour measure.

7. Generally, female students outperformed male students in the Behaviour measure.
8. In general, bilinguals outperformed monolinguals in the Behaviour measure.

Discussion about Bilingual Education

The benefits of bilingualism over monolingualism have been supported in many studies. The benefits have been associated with the enhanced cognition, attention control, and brain functioning in bilinguals. The findings of the present study support a number of previous bilingual studies regarding cognition and attention which have been conducted within the last decade (2003-2013) and support the anecdotal evidence from Aceh that children who were taught English bilingually at state schools outperform children who were taught English monolingually at state schools.

The work of Foy and Mann (2013) revealed that bilingual children outdid monolingual children in nonverbal auditory executive function tasks. Bilingual children significantly outperformed monolingual children on a series of standardised working memory, cognitive control, metalinguistic awareness, and problem-solving ability tests (Lauchlan, Parisi, & Fadda, 2012), and managed cross-linguistic interference more effectively than monolinguals (Bartolotti & Marian, 2012).

Further examples from the recent literature on bilinguals demonstrated higher levels of performance in vocabulary and phonological tasks (Kaushanskaya, Blumenfeld, & Marian, 2011). Bilingual students outperformed monolingual students in reading proficiency (Rauch, Naumann, & Jude, 2011). Bilingual children have advantages over their counterparts in regard to selecting phonological skills when language use and proficiency are controlled (Goldstein & Bunta, 2011). Bilinguals evidenced greater phonological complexity in Spanish than English with respect to word density (Freedman & Barlow, 2011). Bilingual students also scored higher than the monolingual children on a grammatical judgment test (Foursha-Stevenson & Nicoladis, 2011).

Bilingual college students performed better in a task-switching paradigm compared to monolingual college students (Prior & MacWhinney, 2009). Bilinguals were more efficient than monolinguals at inhibiting distracting information (Treccani,

Argyri, Sorace, & Salla, 2009). Bilinguals acquired a newly-learned language faster than monolinguals (Kaushanskaya & Marian, 2009). Bilinguals developed a better ability to maintain action goals to use for bias goal-related information (Colzato et al., 2008). Bilingual participants were faster in performing the task-switching paradigm as well more efficient in the alerting and executive control networks (Costa, Hernandez, & Sebastian-Galles, 2008). Bilingual children maintained better accuracy than monolingual children in dual-task paradigms, especially on visual tasks. Bilingual participants performed better in dual-task letter number category tests (Bialystok, Craik, & Ruocco, 2006). Bilinguals also achieved higher scores than monolinguals in the Simon Task (Bialystok, et al, 2004). Bilinguals had advantages over monolinguals in justifying their evaluation on the appropriateness of certain request strategies in particular contexts as well as on their use of request realisations (Jorda, 2003).

Being bilingualism for Indonesians, more specifically, Acehnese students, is seen as a privilege. There is nothing wrong about learning English, alongside learning Indonesian, despite the ban of bilingual education in Indonesia recently (2013). Learning English is believed to be important because English is now an international language and it is the language of information and technology. Without having ability in English, it is argued that Indonesian students could be predicted to be 'left behind'.

The importance of learning English was mentioned by the first year students of state middle schools in Banda Aceh. Most of them stated that they learnt English because they believed that English could help them go overseas, either for holidays or to pursue their further study. It is convenient for people to travel to neighbouring countries of Indonesia, for example, Singapore, Papua New Guinea, and Australia, if they have a good command in English. They, at least, can read signs at airports and cities or can manage simple conversations with the countries' people, for directions, or simply to get new friends. It is hard to imagine that, if somebody travels overseas without any ability in English, then they would have no trouble understanding signs, attractions, destinations and directions, without some help.

People who want to study overseas are now required to be able to have a good command of English. That requirement is not only for those who want to study at an English-speaking country but, it is also true for students who want to study in a country

which the language is similar to their own. Consider the example of Indonesian students who want to study in Malaysia: even though the Malay language and the Indonesian language are similar, if not identical in some words, and English is not their national language, the students need to pass an English requirement, such as TOEFL and IELTS. The reason behind that is because there are a great number of academic references which are written in English and whose equivalences are scarcely written in or translated into Malay. The case would be true for these students if they want to enroll in any Australian colleges or universities. In this country, English is the language for both academic and daily life. For those who seek a scholarship, their ability in English even is needed to be proved much earlier, that is when they need to satisfy the scholarship panel about their English ability in order that a scholarship be awarded to them.

Some students also mentioned that English could help them get enrolled in their favoured schools and help them get their dream jobs. Now, more favoured schools in Indonesia, especially in metropolitan cities, use English as a medium of instruction. That becomes a trend that attracts candidates of students to enroll at those certain schools. The schools usually run an entry test in order to ensure that only students with a good command of English are selected. Among the selection criteria is the ability for the students to speak, write, and read in English. Students with a low command of English could not go through. Similarly, regarding job vacancies, having a good command in English can give a credit to an employee candidate when competing for a good job. It is more likely that among some employee candidates who have similar academic achievement, age, and ability regarding that specific job, employee candidates with a good command of English would be favoured to get the job.

English is also a language for information and technology. Indonesian students are also aware that a great number of books are written in English. If they want to know the content of the books, they have to know English. In the same line, most technology comes in English. Most of games, smart phones, Facebook, Twitter, and the like, are programmed in English. Most of the manual books and booklets of television, refrigerators, air conditioners, radio cassettes and similar products, are written in English. One may find difficulty in operating them in Indonesia without ability in English.

Based on the discussion above, it establishes that there is nothing wrong about learning English by Indonesians, more specifically, Acehese students. Having more languages is better than having one language. Being bilingual is better than being monolingual. By being bilingual, his or her mind is more powerful than being monolingual. Although bilinguals are not necessarily smarter, they can be more flexible and resourceful. Bilingual learners are better at reasoning, at multitasking, at grasping, and at reconciling conflicting ideas (Kluger, 2013). So it is in the best interests of Indonesian students to be bilingual or multilingual.

Those skills start in the womb, continue at birth, crest at nine months, and decline at six years old. In the uterus, the third trimester is the first time for babies to be able to hear sounds. The first sounds for them to try to recognise are their mother language rhythms. When they are born, the babies recognise their mother tongue and distinguish it from other languages. If their mothers are bilingual, babies recognise both languages. The peak is at nine months when the babies have sharp ears for languages. By one year old, the door to languages begins to close. When the babies are about six years old, they are less natural to languages than they were when they were babies, but they are still better than teenagers or adults (Kluger, 2013).

The question now is whether to be bilingual or multilingual, or be worse off. If one wants to use it, Kluger (2013) suggests that one would better start learning it early, and maintaining it through a lifetime. Maintaining it is not an easy task. Considering a case of two siblings: 11 years and 6 years old who are living in Australia for four years. They have been learning English for four years and at the moment they have reached a resemblance to native Australians. Along the time, they also learn Bahasa Indonesia and Acehese language at home because those two languages are the languages of their parents. Upon returning to Indonesia after these four years, it is reasonably predictable that their ability in Australian English will fade, replaced by Bahasa Indonesia and Acehese language as the languages of the new community. Should they be able to maintain those three languages in a fair share, they might benefit more from the power of the bilingual and multilingual brain. The bilingual and multilingual brain, when activated from birth, can provide advantages for a lifetime, and even can reduce certain cognitive diseases (such as dementia) for old bilinguals and multilinguals to produce a lifetime skill (Kluger, 2013). Figure 12.1 shows bilingualism as a life time skill.

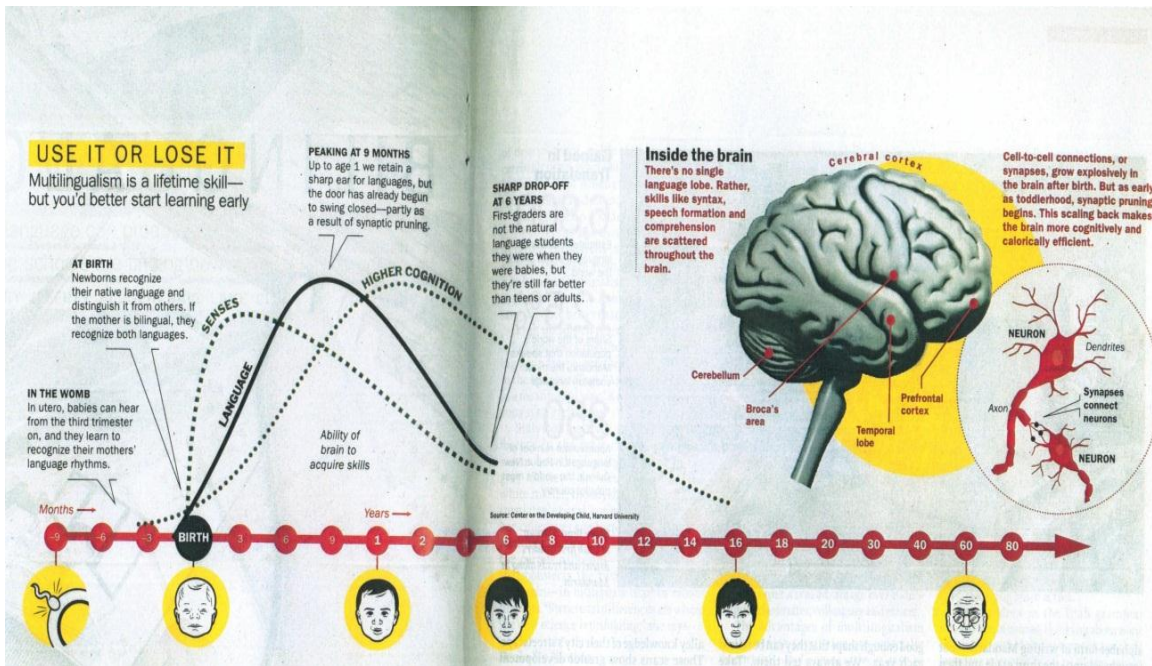


Figure 12.1 Bilingualism as a Life Time Skill

Source: Kluger, J. (2013). The power of the bilingual brain. *Time*, 182(5), pp. 26-27.

The results from Figure 12.1 have important implications for Indonesians in being bilingual or multilingual. It is in the best interests of all Indonesians, in both the long and short term, for bilingual education to be re-instated and to be available to all Indonesian students.

That female students do better at English Comprehension, English Writing and Behaviour with regards to Learning English is probably not a strong concern. In western countries this has been true for many years but males tend to catch up in late secondary school (college) and university. So it is probably not a concern for young Indonesian students as it can reasonably be expected that older male Indonesian students will also catch up in college and university years.

Implications for Policy and Practice

The present study has a number of implications for policy: for the Indonesian Ministry of Education; and for practice, relating to administrators, lecturers, trainee student teachers, Headmasters, ESL/EFL teachers, ESL/EFL students, and parents.

Implications for Policy: Indonesian Ministry of Education/Policy Makers

The implementation of the 2013 English curriculum in Indonesia that banned the bilingual programs from primary level until secondary level has been controversial. Some have seen it as a drawback with regard to learning English in Indonesia. Some have perceived it as a regretful decision, claiming that the decision made to ban bilingual education was immature, lacking in comprehension and lacking in research-based reasons.

The findings of the present study revealed that students who were taught English with two languages performed better than those who were taught English with Bahasa Indonesia only. They also showed that students who were taught bilingually gained better behaviour with regard to learning English. Regarding the implementation of the 2013 curriculum, the findings of the present study could contribute to giving a better understanding to the Indonesian Ministry of Education and its Policy Makers that bilingual education practices could not be blamed for the banning of bilingual education in Indonesia. As there could be a number of other factors with regard to what has been claimed as the 'unfairness of bilingual education in Indonesia', banning it may not be seen as a wise choice. Other factors, like government and school readiness in terms of financial support, human resources, and the like, should be taken into account as well. If the practice of bilingual education in Indonesia was seen as giving a further advantage to those who already have advantages, it should not need to lead to banning it, but to the government implementing bilingual education in all schools. By doing so, the government can maintain bilingual education and its advantages, and at the same time is able to control negative aspects that could harm the practice of bilingual education.

Implications for Practice

Implications for Administrators, Lecturers, and Trainee Student Teachers

The present study showed that teaching English to state middle school students using two languages (English and Bahasa Indonesia) enhanced students' English ability, especially in English Reading Comprehension, and English Writing, as well as improving their behaviour and their perceptions with regard to their learning of English as a foreign language. This has implications for administrators, lecturers, and ESL trainee-student teachers. At the administration level, the study suggested that a well-

planned bilingual education should be prioritised prior to any implementation of bilingual education in Indonesia. The situation in Indonesia tends to show that an ill-planned bilingual education caused the banning of the bilingual program throughout Indonesia. Bilingual programs that were implemented in 2004 were banned and the ban took effect in early 2013. Administrators should base their beliefs on research evidence and not rush to hasty decisions without regard to the evidence.

Previous experience with bilingual education implementation showed Indonesians that bilingual education had been improperly conducted. This led to unsatisfactory results claimed by certain groups who believed that bilingual education in Indonesia created unfairness toward students with low-income and from low socio-economic families (Revianur, 2013b). Further, these certain groups stated that bilingual education could only be enjoyed by students from high-income families who were able to support the schools financially. On the other hand, students from families with low incomes could only manage seats at schools without bilingual programs, creating an unfair perception. Groups of these unsatisfied parents actually brought the case to the Judicial Court (called *Mahkamah Konstitusi* in Indonesia). It resulted in the Judicial Court banning bilingual education in Indonesia from the 2013 (Revianur, 2013b).

However, a better future for bilingual education could be predicted, based on research evidence. The present 2013 English curriculum might last for five years. As Indonesia is dynamic and prone to change, it is highly anticipated that a new curriculum with bilingual programs might be reassigned after five years. If this prediction takes place, it is expected that Indonesian administrators will prepare a new bilingual program prior to its implementation. This is regarded as highly important in order to achieve a 'fairness of Indonesian bilingual education', which is fair and mandatory for each and every student in Indonesia. Therefore, it is hoped that there will be no more banning of bilingual education in the history of Indonesian education.

In line with this, lecturers and teacher trainers at universities should provide a proper and successful transfer of ESL/EFL knowledge, skills, and content to ESL/EFL trainee-student teachers and it should be compulsory, since there are clear benefits to all Indonesians. Trainee-student teachers need to be ready to teach their future students bilingually. One cannot expect well-taught bilingual classrooms if the teachers

themselves are not ready to teach them with knowledge and skills which are changing and demanding (O'Neill, 2008).

Implications for Headmasters, ESL/EFL Teachers, and ESL/EFL Students.

The present study gave evidence that bilingual education enhanced students' ability in English subjects, as it does in many western countries. This information is of importance to headmasters, ESL/EFL teachers, and ESL/EFL students, not only in Aceh Province but also in Indonesia as a whole. Being well-informed about the benefits of bilingual education in Indonesia, headmasters could play crucial roles in managing to prepare for future bilingual education support, in terms of financial support, human resources, administration resources, libraries, and other related school-maintenance support. Headmasters and others need to have access to research reports worldwide showing the benefits of bilingual education for all students and older citizens.

ESL/EFL teachers, in line with this, could try to do their best to implement a bilingual program in their classrooms by conducting teaching and learning processes bilingually, as well as assessing students' work consistently and fairly. They need to constantly upgrade their knowledge and proficiency in English in order to keep them updated with current English practices, techniques, approaches, and skills.

ESL/EFL students can enjoy the learning by using two languages (Bahasa Indonesia and English). Students can succeed in their schools well and acquire high motivation in learning English as a foreign language, but Headmasters need the research evidence to be widely available in order to be able to argue for bilingual education. Students can also use their previous knowledge to help them understand and improve their current knowledge of English. In addition to that, they can benefit from their exposure to English outside their classrooms, as English can also be learnt from books written in English that students read at home or in the library, or from TV programs and music to which they watch or listen, or conversations exchanged with tourists, or on social media like Facebook and Twitter, and the like.

Implication for the Banning of Bilingual Education

The present study could inform the education policy makers about the current reality in Aceh that bilingual education was still needed, apart from the banning of bilingual education throughout Indonesia. It also suggests that the Acehnese government should reconsider the banning of bilingual education in Aceh Province and reestablish bilingual programs at state schools in Aceh with the power it already has as a province. It has the privilege to manage its own education policy, and to learnt from past lessons and previous experience in regard to bilingual education.

Implication for Classroom Practices

The present study also provided implications for success in bilingually-taught classrooms. Teachers of state middle schools in Aceh could take the benefits from this study (and other studies) in order to become well informed about criteria on running a bilingual program, from practices conducted all over the world. The present study also presented examples of well organised English Reading Comprehension and English Writing tests, as well as a good questionnaire to make linear measures. These examples are worth trying in teachers' own classrooms. From some brief informal contacts made by the researcher with some of the respondents, it was revealed that some students have never experienced having Reading Comprehension, English Writing, and Questionnaire tests of the high quality used in this research. The students liked the paper test presentation, due to its clear instructions and interesting pictures that helped them understand the test more easily. This probably means that more money needs to be provided by government and policy makers to enable teachers to have access to high quality material. It also probably means that teachers need better in-service courses that provide research evidence for the benefits of bilingual education.

In practice in Banda Aceh state schools, English test presentations have not always been clear. Instructions have been mostly unclear and always needed further explanation, usually done by the test invigilators on the day of the tests. Pictures, especially colourful pictures, were seldom provided, due to high costs. With the example of the present study's paper test, findings implied that content and presentation of a test are considered equally important, in order that students perform their best.

Implications for Parents

There would seem to be a need to inform parents, through government circles and schools (parent/teacher meetings) about the benefits of bilingual teaching for students with immediate effects for young people and later effects for older people. There are benefits economically and for the young people of Indonesia who are fluent in English and Indonesian. Young Indonesians would be equipped to travel overseas to gain the skills and knowledge that are needed to improve their country. Especially nowadays, the Indonesian government provides a large number of scholarships for eligible candidates to pursue studies overseas, and English mastery is part of the criteria for the selection.

Implications for Further Research

Improved Rasch Measures

In the present study, the researcher designed and constructed novel test formats for three variables: (1) English Reading Comprehension, (2) English Writing, and (3) an Attitude and Behaviour Questionnaire. The three kinds of tests were piloted prior to the main data collection. However, data for two of the three measures (English Reading Comprehension and English Writing) did not show an ideal fit to the Rasch measurement model. This was probably caused by such factors as the difference between student ability and the level of test difficulty and because the students at any one level of ability did not agree about the difficulties of all the items. Future research, therefore, is suggested to examine these issues and perhaps find a model to design (or create) items for English Reading Comprehension and English Writing. It is most probable that a qualitative study would be required where the researcher could talk to and question the students about their views on the test items after the students have answered the items.

It is probable too that, because there would likely be a large difference in ability between pretests and posttests, different tests might be required for the pretests and posttests that could be equated using Rasch measurement, perhaps with some overlapping items.

More Research on Bilingualism and Bilingual Education in Indonesia

There is a scarcity of research on bilingualism and bilingual education in Indonesia and in Aceh province, and it is suggested that more research on these topics should be conducted. It is also important that the research be accessible to Indonesian educators, the Indonesian people and other researchers. There is a possibility that there is some research on bilingualism and bilingual education in Indonesia or in the Aceh province but that is not readily accessible to Indonesians and Indonesian educators.

Inaccessible research and scarcity of research can cause some missing links in regards to future research. This can contribute to poor debating and poor decision-making in Indonesia (as in the recent court case involving the banning of bilingualism in Aceh schools). Therefore, it is important that studies on bilingualism and bilingual education in Indonesia and in Aceh be continuously conducted and the results be made accessible to educators and the public.

Longitudinal Research

Studies on bilingualism and bilingual education involve conducting research over time, so as to allow researchers to make repeated measures that are directly comparable. Longitudinal research is suitable for use in the studies of bilinguals and bilingual education, where the method can investigate developmental trends and comparisons between bilingual students and monolingual students, as well as comparisons among bilingual students themselves over short and long periods of time.

Most longitudinal studies employ observations or measures compared over various time intervals. A large problem here is controlling for extraneous variables and, the longer the time frame, the more likely that extraneous variables (like all the different things that different students do outside school hours), will confound the results. The present study involved a short time frame but there is a clear need to have some longer studies, perhaps over six or twelve months. However, this can be costly and difficult to manage. It is probable that a large amount of joint university researcher and teacher cooperation, and perhaps parent cooperation, will be needed. Also, pretest and posttest equating may be needed using the latest Rasch measurement techniques to link pretests with posttests that may contain different items.

Teachers and parents

Studies on bilingualism and bilingual education can cover a wide range of people because many people (teachers, administrators, parents and policy-makers) are involved in children's education. Teachers and parents are believed to have the most influence and power in teaching bilingualism and in promoting academic achievement. Research on teachers' and parents' influences in relation to students' bilingualism and bilingual education in Aceh is inadequate. So, future research should consider teachers and parents as the focus of the studies on bilingualism and bilingual education in Indonesia and Aceh province. This could involve a variety of approaches including control and experimental groups with pretest and posttest Rasch created linear measures, longitudinal studies and qualitative studies. As already mentioned, controlling for extraneous variables will be a problem and researchers will probably have to use a mixture of physical and statistical controls with more accurate Rasch measures to be able to make better conclusions from the data.

English listening and speaking skills

The present study focused on two English skills: Reading Comprehension and Writing but it did not investigate Listening and Speaking skills. Future research is encouraged to investigate students' English abilities in listening and speaking because these are also important English skills. Future research results could be combined with the present study's findings (repeated with improved Rasch measures) in various longitudinal and qualitative studies in Aceh. Altogether, a better view would be gained of students' abilities and achievement in the four skills related to learning English as a foreign language in the Acehnese context.

Limitations of the Study

This study, inevitably, had some shortcomings, in terms of a number of aspects:

Managing Time and Extraneous issues

The experimental study was conducted within two months, which is equivalent to 16 meetings. This short duration was due to three reasons: (1) to allow sufficient meetings for students to experience the two medium of instruction (bilingual mode and

monolingual mode), (2) to avoid inconvenience for teachers, and (3) to manage the control of extraneous variables.

The duration of 16 meetings was predicted to be sufficient for students to be taught either bilingually or monolingually in order to allow them to experience their English learning progress. It was also expected that at the end of the period, the students would be ready to sit for the posttests with expected changes in terms of their ability in English Reading Comprehension, English Writing, and the Attitude/Behaviour Questionnaire with regard to their learning English as a foreign language. The duration of two months was also seen as a convenient time allowance for the teachers to participate. It was strongly predicted that beyond two months, teachers might feel disturbed due to their other responsibilities and commitments to the schools and family. The duration was also limited in order to control for extraneous variables related to both types of students involved in the study. Both bilinguals and monolinguals needed to maintain similarities in a number of aspects, such as the same English lesson content, the same number of lessons and lesson durations, the same pretest and posttest duration, the same homework, and the same text books, but the nature of the bilingual and monolingual teaching program needed to be left unchanged. The only difference between the two groups that the study intended to maintain was that the one group experienced bilingual teaching and learning processes, while the other group experienced monolingual teaching and learning processes. This similarity needed to be maintained in order that any changes that appeared after the time frame of two months could be associated with the change due to the type of teaching instruction.

The changes, measured in this current study, therefore, suggest that bilinguals were better than monolinguals in the three tests: English Reading Comprehension, English Writing and Behaviour Questionnaire, due to the fact that bilinguals were taught using both English and Bahasa Indonesia, while monolinguals were taught using only Bahasa Indonesia as the medium of instruction.

However, the present study would be more reliable if it extended the experimental duration, that is, more than two months, providing extraneous variables could be sufficiently well controlled. This assumes that the longer duration between pretest and protest would show that students could demonstrate better ability in the

posttest. This improved ability could be associated with the use of English and Bahasa Indonesia (bilingually) in teaching English as a subject in the Acehese context. The study would be more comprehensive if it were conducted over three school years, starting at year 7 and concluding at year 9.

Managing Teacher Perceptions (through Interviews)

The present study was conducted with quantitative data, with a small amount on qualitative data. For future research it is suggested that, rather than just students, the study also collect data on the perceptions of teachers. Teachers' perceptions could enrich our understanding of their experiences and expectations regarding their own way of teaching English, as well as their assumptions and hopes for their students with regard to learning English as a foreign language.

Managing Classroom Observations

The present study did not investigate the way a classroom of teaching and learning was run, either. Therefore, it is recommended that future research could examine classroom observations of bilingual and monolingual teaching. Classroom observation data offer a valuable understanding and knowledge of how bilingual students interact in a classroom, compared to monolinguals, during the teaching and learning process.

Measures were not Ideal and Need Improvement

The present study used data collected from 780 students in 18 state middle Schools in Banda Aceh, with both males and females, and bilinguals and monolinguals. The test designs for the three measures (English Reading Comprehension, English Writing, and Behaviour Questionnaire) were especially constructed by the researcher due to the fact that there were no existing tests that were suitable for the present study. The three tests were piloted prior to collecting the main data using a class of bilinguals and a class of monolinguals. Because True Score Theory of Measurement (like percentage scores) – the measurement method universally used in schools – is non-linear, a modern method of measurement – Rasch measurement – was used in the present study to create linear and more reliable measures. While these Rasch measures were better than the percentage scores used by teachers, they were not as good as the

researcher would have liked because the fit to the measurement model was not ideal. However, many significant results were still able to be obtained. For future use of these tests, it is recommended that the tests be modified and amended.

There is a need for high quality tests of English writing and English comprehension that would be applicable to many different standards of English and so much further research is needed to produce them, preferably using Rasch measurement techniques. Alternatively, there is a need for a theoretical method that enables researchers to design tests from first principles that will produce data that fit a Rasch measurement model without any misfit. That could well be a better outcome, if someone could create such a model.

REFERENCES

- Abu Rabia, S., & Siegel, L. S. (2011). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of Psycholinguistic Research*, 31(6), 661-678.
- ADB, A. D. B. (2009). Aceh-Nias rehabilitation and reconstruction, from <http://www.adb.org/Documents/Produced-Under-TA/39127/39127-01-INO-DPTA.pdf>.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlation of bilingualism. *Review of Educational Research*, 80(2), 207-245.
- Adi-Japha, E., Berberich-Artzi, J., & Lidnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development Psychology*, 46, 93-105.
- Alarcon, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350.
- Ali, Z., Mukundan, J., Baki, R., & Ayub, A. F. M. (2012). Second language learners' attitude towards the methods of learning vocabulary. *English Language Teaching*, 5(4), 24-36.
- Alshumaimeri, Y. A., & Almasri, M. M. (2012). The effect of using WebQuest on reading comprehension performance of Saudi EFL students. *The Turkish Online Journal of Educational Technology*, 11(4), 295-306.
- Andrich, D. (1988a). *Rasch models for measurement*. Newbury Park, CA: Sage.
- Andrich, D. (1988b). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, 1(4), 363-378.
- Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath & S. H. Lovibond (Eds.), *Mathematical and theoretical systems* (pp. 7-16). North Holland: Elsevier Science.
- Andrich, D., Sheridan, B. E., & Luo, G. (2010). *RUMM2030: A windows-based item analysis program employing Rasch unidimensional measurement models*. Perth, Western Australia: RUMM Laboratory.
- Andrich, D., & van-Schoubroeck, L. (1989). The general health questionnaire: A psychometric analysis using latent trait theory. *Psychological Medicine*, 19, 469-485.

- Answer.Com. (2010). Aceh. Retrieved from <http://www.answers.com/topic/aceh>.
- Arshad, M., Ali, M. & Naseem, C. (2013). Does gender difference matter in learning English in a culture? *Interdisciplinary Journal of Contemporary Research in Business*, 5(3), 409-419.
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189-208.
- Baetens-Beardsmore, H. B. (2003). Who's afraid of bilingualism? In J.-M. Dewaele, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles*. (pp. 10-27). Clevedon, U. K.: Multilingual Matters.
- Baker, C. (1996). *Foundations of bilingual education and bilingualism* (2nd ed.). Clevedon, UK: Multilingual Matters.
- Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Bristol: UK: Multilingual Matters.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Bristol: Multilingual Matters.
- Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, 36(2), 1129-1147.
- BBC. (2005). Indonesia. Retrieved from http://news.bbc.co.uk/2/shared/spl/hi/world/04/asia_quake/quake_maps/html/1.stm.
- Bergman, M. M. (2008). *Advances in mixed-methods research: Theories and applications*. Thousand Oaks, C A: Sage.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13(2), 116-137.
- Bialystok, E. (2010). Global-local and Trail-Making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology*, 46(10), 93-105.
- Bialystok, E. (2011). Reshaping the mind: The benefit of bilingualism. *Journal of Experimental Psychology*, 65(4), 229-235.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122(1), 67-73. doi: 10.1016/j.cognition.2011.08.003
- Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptom of dementia. *Neuropsychologia*, 45, 459-464.

- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.
- Bialystok, E., Craik, F. I. M., & Ruocco, A. (2006). Dual-modality monitoring in a classification task: The effects of bilingualism and ageing. *The Quarterly Journal of Experimental Psychology, 59*(11), 1968-1983.
- Bialystok, E., & Herman, J. (1999). Does bilingualism matter for early literacy. *Bilingualism: Language and Cognition, 2*(1), 35-44.
- Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages bilingual children in two cultures. *Cognition, 112*, 494-500.
- Biedron, A., & Szczepaniak, A. (2012). Working memory and short-term memory abilities in accomplished multilinguals. *The Modern Language Journal, 96*, 290-306.
- Bolukbas, F., Keskin, F., & Polat, M. (2011). The effectiveness of cooperative learning on the reading comprehension skills in Turkish as a foreign language. *TOJET: The Turkish Online Journal of Educational Technology, 10*(4), 330-335.
- Brisk, M. E. (2005). Bilingual education. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97-113.
- Butler, Y. G. (2013). Bilingualism/multilingualism and second language acquisition. In T. K. Bathia & W. C. Ritchie (Eds.), *Handbook of bilingualism and multilingualism* (2nd ed., pp. 109-136). Somerset, NJ: Wiley.
- Buyuktaskapu, S. (2012). Effect of family supported pre-reading training program given to children in preschool education period on reading success in primary school. *Educational Sciences: Theory & Practice, 12*(1), 309-316.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282-298.
- Cazden, C. B., & Snow, C. E. (Eds.). (1990). *English plus: Issues in bilingual education*. London: Sage.
- Chapman, E. S. (2006). True score. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 1, pp. 421-423). Thousand Oaks, CA: Sage.
- Chodidjah, I. (2009). *Structured and monitored teacher development: The wind of change*. Paper presented at the the access English EBE symposium, Jakarta.

- Chow, B. W.-Y., McBride, C., Cheung, H., & Chow, C. S.-L. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology, 44*, 233-244.
- Clark, V. P., & Creswell, V. L. P. (2008). *The mixed methods reader*. Thousand Oaks, C A: Sage.
- Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. B. (2006). Prevalence of mixed methods sampling designs in social science research. *Evaluation and Research in Education, 19*(2), 83-101.
- Colome, A. (2001). Lexical activation in bilinguals' speech production: Language specific or language-independent? *Journal of Memory and Language, 45*(4), 721-736.
- Colzato, L. S., Bajo, M. T., Wildenberg, W. v. d., Paolieri, D., Nieuwenhuis, S., Heij, W. L., et al. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Recognition, 34*(2), 302-312.
- Conway, A. R. A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variation in working memory*. New York: Oxford University Press.
- Costa, A., Hernandez, M., & Sebastian-Galles, N. (2008). Bilingualism and conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59-86.
- Creswell, V. L. P., & Clark, V. P. (2006). *Designing and conducting mixed methods research*. Thousand Oaks, C A: Sage.
- Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish-English bilingual. *Applied Linguistics, 20*, 1-20.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
- Cummins, J. (2003). Bilingual education: Basic principles. In J.-M. Daelewe, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles. Festschrift in honour of Hugo Baetens Beardsmore* (pp. 56-66). Clevedon, England: Multilingual Matters.
- Davies, R. (2005, October 9). Rising to the challenge of bilingual education, online, *The Jakarta Post*. Retrieved from <http://www.thejakartapost.com/news/2005/07/16/bilingual-education-often-misunderstood-always-complicated.html>

- Depdiknas, N. E. D. (2007). *Pedoman penjaminan mutu sekolah/madrasah bertaraf internasional jenjang pendidikan dasar dan menengah*. Jakarta: Direktorat Tenaga Kependidikan, Direktorat Jenderal Peningkatan Mutu Pendidikan dan Tenaga Kependidikan, Departemen Pendidikan Nasional.
- Denscombe, M. (2010). *The good research guide: for a small-scale social research project* (4th ed.). Berkshire, UK: Open University Press.
- Diamond, A. (2002). Normal development of pre-frontal from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Strauss & R. Knight (Eds.), *Principle of frontal lobe function* (pp. 466-503). New York: Oxford.
- Dicker, S. J. (Ed.). (2003). *Languages in America: A pluralistic view* (2nd ed.). Clevedon: Multilingual Matters.
- Dixon, L. Q., Zhao, J., Shin, J.-Y., Wu, S., Su, J.-H., Burgess-Brigham, R., et al. (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, 82(5), 5-60.
- Education, G. (2010). Indian ocean tsunami. Retrieved from <http://www.globaleducation.edna.edu.au/globaled/go/pid/2258>
- Edwards, J. R. (2010). *Minority languages and group identity: Cases and categories*. Amsterdam: John Benjamins.
- Elvyanti, S. (n.d). Indonesian curriculum history. Retrieved from <http://file.upi.edu/Direktori/E%20-%20FPTK/JUR.%20PEND.%20TEKNIK%20ELEKTRO/197311222001122%20-%20SISCKA%20ELVYANTI/sejarah%20kurikulum1.pdf>
- Exchange., S. (n.d). The education system in Indonesia. Retrieved from <http://www.spainexchange.com/guide/ID-education.htm>
- Ferguson, C. A., Houghton, C., & Wells, M. H. (1977). Bilingual education: An international prespective. In B. Spolsky & R. Cooper (Eds.), *Frontier of bilingual education*. Rowley, MA: Newbury House.
- Forum, A. R. (2006). Mendiknas resmikan "Fatih Bilingual School". Retrieved from <http://www.acehrecoveryforum.org/id/index.php?action=PSCM&no=2437>
- Foursha-Stevenson, C., & Nicoladis, E. (2011). Early emergence of syntactic awareness and cross-linguistic influence in bilingual children' judgments. *International Journal of Bilingualism*, 15(4), 521-534.
- Foy, J. G., & Mann, V. A. (2013). Bilingual children show advantages in nonverbal auditory executive function task. *International Journal of Bilingualism*, 0(0), 1-13.

- Francis, N. (1999). Bilingualism, writing, and metalinguistic awareness: Oral-literate interaction between first and second languages. *Applied Psycholinguistics*, 20, 533-561.
- Freeman, R. (2007). Reviewing the research on language education programs. In O. Garcia & C. Baker (Eds.), *Bilingual education: An introductory reader*. Clevedon: Multilingual Matters (pp. 3-18).
- Freedman, S. E., & Burlow, J. (2011). Using whole-word production measures to determine the influence of phonotactic probability and neighborhood density on bilingual speech production. *International Journal of Bilingualism*, 16(4), 369-387.
- Gebhard, J. G. (2012). International students' adjustment problems and behaviors. *Journal of International Students*, 2(2), 184-193.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Rowley, M A: Newbury House.
- Ghonsooly, B., & Shawqi, S. (2012). The effects of foreign language learning on creativity. *English Language Teaching*, 5(4), 161-167.
- Goldstein, B. A., & Bunta, F. (2011). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, 16(4), 388-401.
- Government, A. (2000). Peraturan Daerah Provinsi Daerah Istimewa Aceh Nomor 4 Tahun 2000. Retrieved from http://www.acehprov.go.id/images/stories/file/Perda/perda_prov_nad_no_04_tahun_2000.pdf
- Grech, H., & McLeod, S. (2012). Multilingual speech and language and disorder. In D. Battle (Ed.), *Communication disorders and development in multicultural population* (4th ed., pp. 120-147). St. Louis, MO: Elsevier.
- Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, 28, 41-50.
- Greene, J. C., Caracelli, V. J., & Graham, W. E. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Holland, P. W., & Wainer, H. (1993) *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum
- Hornberger, N. H. (1991). Extending enrichment bilingual education: Revisiting typologies and redirecting policy. In O. Garcia (Ed.), *Bilingual education*

focusschrisft in honour of Joshua A. Fisherman (Vol. 1, pp. 125-234).
Philadelphia: John Benjamins.

- Hosseini, S. B. (2012). Asynchronous computer-mediated corrective feedback and the correct use of prepositions: Is it really effective? *Turkish online Journal of Distance Education, TOJDE*, 13(4), 95-111.
- Jafari, N. & Ansari, D. N. (2012). The effect of collaboration on Iranian EFL learners' writing ability. *International Studies*, 5(2), 125-131.
- Jallad, N.Y. & Bani Abdelrahman, A. A. (2008). The Effect of Multiple Intelligences Strategies on EFL Ninth Graders' Achievement in Reading Comprehension. *Indian Journal Of Applied Linguistics*, 34(1-2), 87.
- Jones, D. V., & Martin-Jones, M. (2004). Bilingual education and language revitalisation in Wales: Post achievement and current issues. In J. W. Tollefson & A. B. M. Tsui (Eds.), *Medium instruction policies: Which agenda? Whose Agenda?* Mahwah, NY: Erlbaum.
- Jorda, M. P. S. (2003). Metapragmatic awareness and pragmatic production of third language learners of English: A focus on request acts realizations. *International Journal of Bilingualism*, 7(1), 43-69.
- Judge, J. W. (2011). Use of language learning strategies by Spanish adults for business English. *International Journal of English Studies*, 12(1), 705-710.
- Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2011). The relationship between vocabulary and short-term memory measures in monolingual and bilingual speakers. *International Journal of Bilingualism*, 15(4), 408-425.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin Review*, 16(4), 705-710.
- Kave, G., Eyal, N., Shorek, A., & Cohen-Manfield, J. (2008). Multilingualism and cognitive state in the oldest old. *psychology and aging*, 23(1), 70-78.
- Keats, J. A. (1997). Classical test theory. In J. P. Keeves (Ed.), *Educational research, methodology and measurement* (2nd ed., pp. 713-719). Cambridge, U K: Pergamon Press.
- Kemp, C. (2007). Strategic processing in grammar learning: Do multilinguals use more strategies? *International Journal of Bilingualism*, 4, 241-261.
- Kharkhurin, A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilingual divergent thinking. *Bilingualism: Language and Cognition*, 11(2), 225-243.

- Kharkhurin, A. V. (2009). The role of bilingualism in creative performance on divergent thinking and invented alien creatures texts. *Bilingualism and Creativity*, 43(1), 59-71.
- Kieffer, M. J., & Lesaux, N. K. (2012). Development of morphological awareness and vocabulary knowledge in Spanish-speaking language minority learners: A parallel process latent growth curve model. *Applied Psycholinguistics*, 33(1), 23-54.
- Kirkpatrick, A. (2007). *World englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.
- Kirkpatrick, A. (2012). English in ASEAN: Implications for regional multilingualism. *Journal of Multilingual and Multicultural Development*, 33(4), 331-344.
- Kirkpatrick, A., & Sussex, R. (Eds.). (2012). *English as an international language in Asia: Implications for language education*. London, UK: Springer.
- Kluger, J. (2013). The power of the bilingual brain. *Time*, 182, 24-29.
- Knoch, U. (2010). Investigating the effectiveness of individualized feedback to rating behavior: A longitudinal study. *Language Testing*, 28(2), 179-200.
- Koh, K. H. (2008). *Using Rasch measurement to construct a diagnostic reading assessment battery*. Paper presented at the the third International Rasch conference, University of Western Australia, Perth.
- Kovacs, A. M., & Mehler, J. (2009). *Cognitive gains in 7-month old bilingual infants*. Paper presented at the the national academy of Sciences of the united states of America.
- Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, but not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism* 9(2), 119-135.
- Kroll, J. F., & deGroot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. D. Groot & J. F. Kroll (Eds.), *Tutorial in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Lauchlan, F., Parisi, M., & Fadda, R. (2012). Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a 'minority' language. *International Journal of Bilingualism*, 17(1), 43-56.
- Laurent, A., & Martinot, C. (2009). Bilingualism and phonological segmentation of speech: The case of English-French pre-schoolers. *Journal of Early Childhood Literacy*, 9, 29-49.

- Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review*, 63, 605-628.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543-560.
- Lin, A. M. Y., & Man, E. Y. F. (2009). *Bilingual education: Southeast Asia perspectives*. Hong Kong: Hong Kong University Press.
- Mackey, W. F. (1978). The importation of bilingual education models. In J. Alatis (Ed.), *Georgetown University Roundtable: International dimensions of education*. Washington DC: Georgetown University Press.
- Maher, J. (1997). Linguistic minorities and education in Japan. *Educational Review*, 49(2), 115-127.
- Maley, C., & Bond, T. G. (2011). Measuring cognitive development in early childhood learning environments. In R. F. Cavanaugh & R. F. Waugh (Eds.), *Application of Rasch measurement in learning environments research*. Rotterdam: Sense Publishers.
- Marashi, H., & Dadari, L. (2012). The impact of using task-based writing on EFL learners' writing on EFL learners' writing performance and creativity. *Theory and Practice in Language Studies*, 2(12), 2500-2507.
- Marian, V., Spivey, M., & Hirsh, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86(1), 70-82.
- Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4), 279-298.
- Masters, G. N. (1997). Partial credit model. In J. P. Keeves (Ed.), *Educational research, methodology and measurement* (2nd ed., pp. 857-863). Cambridge, U K: Pergamon Press.
- McMurray, A. J., Pace, R. W., & Scott, D. (2004). *Research: A common sense approach*. Victoria, W A: Thomson.
- Mekheimer, M. A. A. (2012). Assessing aptitude and attitude development in a translation skills course. *CALICO Journal*, 29(2), 321-340.
- Menanet. (n.d). National education. Retrieved from <http://home.menanet.net/~indoemb/indoemb/univ.htm>.
- Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, 23, 226-235.

- Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: multilingualism and multilingual education* (pp. 35-63). Clevedon, U K: Multilingual Matters.
- Metsämuuronen, J., Svedlin, R., & Ilic, J. (2012). Change in pupils' and students' attitude towards school as a function of age: A Finnish perspective. *Journal of Educational and Development Psychology*, 2(2), 134-151.
- Michel, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, N J: Lawrence Erlbaum.
- Michel, J. (1999). *Measurement in psychology: A critical history of methodological concept*. Cambridge, U K: Cambridge University Press.
- Miles, A., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, C A: Sage.
- Miyake, A., Friedman, N. P., Emerson, M. J., & Witzki, A. H. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A talent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Mohandas, R. (n.d). Overview of education for sustainable development (ESD) in Indonesia. Retrieved from www.ias.unu.edu
- Morales, J., Calvo, A., & Bialystok, A. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114, 187-202.
- Morse, M. M., & Niehaus, L. (2009). *Mixed method design: principles and procedures*. Walnut Creek, California: Left Coast.
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantages. *Developmental Science*, 10(6), 179-726.
- Murad Sani, A. & Zain, Z. (2011). Relating adolescents' second language reading attitudes, self-efficacy for reading, and reading ability in a non-supportive ESL setting. *Reading Matrix: An International Online Journal*, 11(3), 243-254.
- Nakanishi, T., & Ueda, A. (2011). Extensive reading and the effect of shadowing. *Reading in a Foreign Language*, 23(1), 1-16.
- Nasional, D. P. (2001). Kebijakan umum pendidikan dasar dan menengah. Retrieved from <http://sibi.or.id/download/Bing.doc>
- Nurdin, M. (2006). Banda Aceh overview after one year Tsunami. *Journal of Disaster Research*, 1(1), 116-117.
- O'Neill, S. (2008). A case study of learning in English as a foreign language (EFL) in Japan: High school students' English proficiency levels and fostering positive

cross-cultural attitudes. *International Journal of Pedagogy and Learning*, 4(5), 104-120.

Ovando, C. J., & Combs, M. C. (2012). *Bilingual and ESL Classrooms: Teaching in multicultural contexts* (5th ed.). New York: McGraw-Hill.

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the hospital anxiety and depression scale (HADS). *British Journal of Clinical Psychology*, 46, 1-18.

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effect of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, 108(3), 567-579.

Prior, A., & MacWhinney, B. (2009). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253-262.

Rahmani, M., & Sadeghi, K. (2011). Effects of note-taking training on reading comprehension and recall. *Reading Matrix: An International Online Journal*, 11(2), 116-129.

Ransdell, S., Barbier, M.-L., & Niit, T. (2006). Metacognitions about language skills and working memory among monolingual and bilingual college students: When does multilingualism matter? *The International Journal of Bilingual Education and Bilingualism*, 9, 728-741.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark's Paedagogiske Institut.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (expanded edition, originally published in 1960). Chicago, IL: Mesa Press.

Rasch, G. (2010). *Probabilistic models for some intelligence and attainment tests (50th anniversary ed.)*. Chicago, IL: Mesa Press.

Rauch, D. P., Naumann, J., & Jude, N. (2011). Metalinguistic awareness effects of full biliteracy on third-language reading proficiency in Turkish-German bilinguals. *International Journal of Bilingualism*, 16(4), 402-418.

Revianur, A. (2013a, 8 January). RSBI dan SBI jadi sekolah biasa, online, *Kompas.com*. Retrieved from http://edukasi.kompas.com/read/2013/01/08/18233422/RSBI.dan.SBI.Jadi.Sekolah.Biasa?utm_source=edukasi&utm_medium=cpc&utm_campaign=artbox

Revianur, A. (2013b, 8 January). Ini alasan MK batalkan status RSBI/SBI, online, *Kompas.com*. Retrieved from <http://edukasi.kompas.com/read/2013/01/08/18431250/Ini.Alasan.MK.Batalkan.Status.RSBISBI#komentar>

- Rispen, J. E., McBride-Chang, C., & Reitsma, P. (2008). Morphological awareness and early and advanced word recognition and spelling in Dutch. *Reading and Writing, 21*, 587-607.
- Rocco, T. S., Bliss, L. A., Gallagher, S., Perez-Prado, A., Alacaci, C., Dwyer, E. S., et al. (2003). The pragmatic and dialectical lenses: Two views of mixed methods use in education. In A. Tashakkori & C. Teddlie (Eds.), *The handbook of mixed methods in the social and behavioural sciences*. Thousand Oaks, CA: Sage.
- Rodriquez-Fornells, A., Balaguer, R. D., & Munte, T. F. (2006). Executive control in bilingual language processing. *Language Learning, 56*(1), 133-190.
- Rodriquez-Fornells, A., Rotte, M., Heinze, H. J., Nosselt, T., & Munte, T. F. (2002). Brain potential and functioning MRI evidence for how to handle two languages with one brain. *Nature, 415*, 1026-1029.
- Rossell, C. H., & Baker, K. (1996). The effectiveness of bilingual education. *Resources in the Teaching of English, 30*(1), 7-74.
- Roy, S., & Galiev, A. (2011). Discourses on bilingualism in Canadian French immersion programs *Canadian Modern Language Review, 67*(3), 351-376.
- RUMM 2030 Manual*. (2009). Perth, Western Australia: RUMM Laboratory.
- Sadeghi, K. & Sharifi, F. (2013). The effect of post-teaching activity type on vocabulary learning of elementary EFL learners (2013). *English Language Teaching, 6*(11), 65-76.
- Sahin, A. (2011). Effects of linear texts in page scrolling and page-by-page reading forms on reading comprehension introduction. *TOJET: The Turkish Online Journal of Educational Technology, 10*(3), 94-100.
- Shih, S. (2012). *Bilingualism in school and society: Language, identity, and policy*. New York: Taylor and Francis.
- Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement *Structural Equation Modelling* (Vol. 3, pp. 25-40).
- Soleimani, H., Mainnzadeh, A., Kassaian, Z., & Ketabi, S. (2012). The effect of instruction based on multiple intelligence theory on the attitude and learning of general English. *English Language Teaching, 5*(9), 45-53.
- Sumiya, H., & Healy, A. (2004). Phonology in the bilingual stroop effect. *Memory and Cognition, 32*, 752-758.
- Swanson, H. L., Orosco, M. J., Lussier, C. M., Gerber, M. M., & Guzman-Orth, D. A. (2011). The influence of working memory and phonological processing on English language learner children's bilingual reading and language acquisition. *Journal of Educational Psychology, 103*(4), 838-859.

- Syahdan, S. (2009). Developing characteristics of excellent teachers towards international standardised school. Retrieved from <http://pps.unnes.ac.id/pps1/files/paper%20seminar%20pbi%202009/Syahdan.doc>
- Tabatabaei, O., & Assefi, F. (2012). The effect of portfolio assessment technique on writing performance of EFL learners. *English Language Teaching*, 5(5), 138-147.
- Tajeddin, Z., & Daraee, D. (2013). Vocabulary acquisition through written input: Effects of form-focused, message oriented, and comprehension tasks. *Teaching English as a Second or Foreign Language*, 16(4), 1-19.
- Treccani, B., Argyri, E., Sorace, A., & Salla, S. D. (2009). Spatial negative priming in Bilingualism. *psychonomic Bulletin & Review*, 16(2), 320-327.
- Tubagus, A. (2005). Foreign teachers not qualified. *Asia Africa Intelligence Wire*. Retrieved from <http://www.accessmylibrary.com/article-1G1-136777930/foreign-teachers-not-qualified.html>
- Vorstman, E. L. P., Swart, H. d., Ceginskas, V., & Bergh, H. v. d. (2009). Language learning experience in school context and metacognitive awareness of multilingual children. *International Journal of Multilingualism*, 6(3), 258-280.
- Wang, C., Kim, D.-H., Bong, M., & Ahn, H. S. (2012). Examining measurement properties of an English self-efficacy scale for English language learners in Korea. *International of Educational Research*, 59, 24-34.
- Washington, C. (2010). Aceh. Retrieved from <http://courses.washington.edu/larescue/aceh6.htm>
- Waugh, R. F. (2007). Rasch measurement. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 820-825). Thousand Oaks, CA: Sage.
- Waugh, R. F. (Ed.). (2003). *On the forefront of educational psychology*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2005). *Frontiers in educational psychology*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2010a). *Specialized Rasch measures applied at the forefront education*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2010b). *Applications of Rasch measurement in education*. New York: Nova Science Publishers.
- Waugh, R. F., Bowering, M., & Chayarathee, S. (2005). Creating scales to measure reading comprehension, and attitude and behaviour, for prathom 6 (grade 6) students taught ESL through a cooperative learning method in Thailand. In R. F.

Waugh (Ed.), *Frontiers in Education Psychology* (pp. 183-219). New York: Nova Science Publishers.

Waugh, R. F., Bowering, M. H., & Torok, S. (2005). Creating scales to measure reading comprehension, and attitude and behaviour, for prathom (grade) 7 students taught ESL through a genre-based method in Thailand. In R. F. Waugh (Ed.), *Frontiers in Education Psychology* (pp. 133-173). New York: Nova Science Publishers.

Waugh, R. F., & Chapman, E. (2005). An analysis of dimensionality using factor analysis (true score theory) and Rasch measurement: What is the difference? Which method is better. *Journal of Applied Measurement*, 6(1), 80-99.

Wei, L. (2000). *The bilingualism reader*. London: Routledge.

Wei, L. (2010). BAMFLA: Issues, methods and directions. *International Journal of Bilingualism*, 14(1), 3-9.

Wijaya, H. P. S. (2008). The impacts of changes in education policy on ELT in Indonesia. Retrieved from [http://fportfolio.petra.ac.id/user_files/04-036/The%20Impacts%20\(full%20complete\).doc](http://fportfolio.petra.ac.id/user_files/04-036/The%20Impacts%20(full%20complete).doc)

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modelling*, 3(1), 3-24.

Wu, Y. J., & Thierry, G. (2010). Investigating bilingual processing: The neglected role of language context. *Frontiers in Language Sciences*, 1, 1-6.

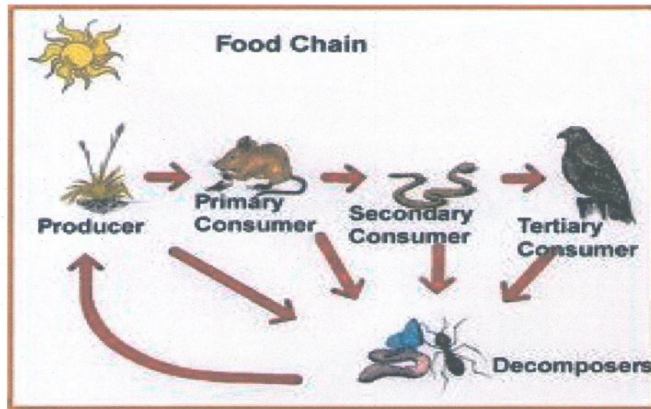
Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to difference in reading behaviour, text comprehension and noticing of language form? *Language Teaching Research*, 10(4), 419-434.

Yuwono, G. (2005). *English language teaching in decentralised Indonesia: Voices from the less privileged schools*. Paper presented at the AARE 2005 International Education Research Conference. <http://www.aare.edu.au/05pap/yuw05050pdf>

APPENDICES

Appendix A-English Reading Comprehension

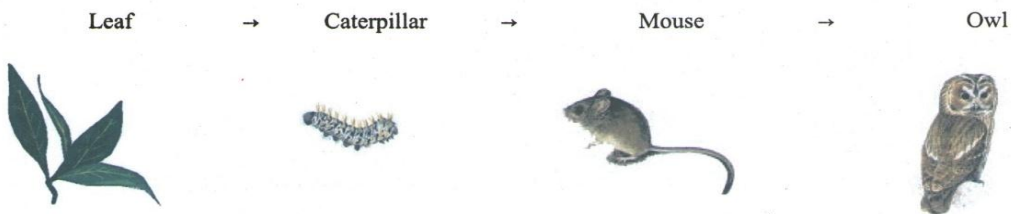
FOOD CHAINS



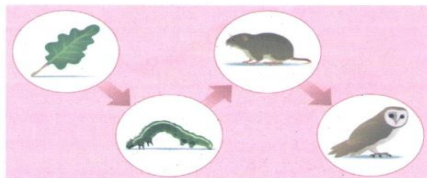
NEEDS FOR ENERGY

When you need energy in the morning before you come to school what do you do? Eat breakfast. Living things need food to give them energy. Plants get energy from sunlight. Caterpillars get their energy from plants. Mice get their energy from caterpillars. Owls get their energy from mice. Plants, caterpillars, mice and owls when die, give energy to decomposers. The plants and animals in a food chain depend on each other to live. What would happen if all the plants died? The caterpillars would die. All the animals that eat depend on the caterpillars would die too. So every part of the chain is important.

Look at this picture representing a food chain using arrows,



Or this picture showing links or chains.



The caterpillar eats a leaf

The mouse eats the caterpillar

The owl eats the mouse

These links between animals and plants are called **food chains**

The food chain consists of four main parts: the sunlight, producers, consumers, and decomposers.

SUNLIGHT AND PRODUCERS

The Sunlight. All food chains begin with the sunlight. Plants use the sunlight to grow.

Producers: Plants are called producers because they make their own food and feed herbivores.

CONSUMERS AND DECOMPOSERS

Consumers: Animals are called consumers because they eat other plants and animals. Consumers are called differently according to what they eat.

1. Some animals eat plants, they are called herbivores. Herbivores are also called **primary consumers**.
2. Some animals eat other animals, they are called carnivores. Carnivores are also called **secondary consumers**.
3. Some animals eat and harm other animals, they are called parasites.
4. Some animal eat dead animals, they are called scavengers.

Decomposers: When plants and animal died, some bacteria and fungi change them into gases such as carbon and nitrogen. These gases are then released back into the air, soil, or water.

HUMANS IN THE FOOD CHAINS

Human are called omnivores because they can eat plants and animals. Humans have superior brains so they often eat animals that are larger than themselves.

[Adopted from various sources]

DIRECTIONS:

1. Put **your name** on the answer sheet provided.
2. Read the passage carefully and cross (X) a, b, or c that suits your answer best.

QUESTIONS:

Needs for Energy

1. What will happen if there is no food?
 - a. No problem
 - b. Plants will die
 - c. Plants and animals will die
2. From where do plants get their energy?
 - a. Animals
 - b. Sunlight
 - c. Humans

3. A food chain is
 - a. A link between plants and plants
 - b. A link between plants and animals
 - c. A link between animals and animals

Sunlight and Producers

4. What do plants use to grow?
 - a. Animals
 - b. Sunlight
 - c. Humans
5. What do we call plants?
 - a. Decomposers
 - b. Consumers
 - c. Producers
6. Plants make
 - a. their own food
 - b. other animals' food
 - c. their own food and other animals' food

Consumers and decomposers

7. What are primary consumers?
 - a. Animals that eat plants
 - b. Animals that eat plants and other animals
 - c. Animals that eat other animals
8. What is an example of a carnivore?
 - a. Cow
 - b. Goat
 - c. Lion
9. Decomposers are
 - a. Goats that eat grass
 - b. Bacteria and fungi that eat dead things
 - c. Cats that eat fish

Humans in the Food Chains

10. What do humans eat?
 - a. Plants
 - b. Animals
 - c. Plants and animals
11. What do we call humans?
 - a. Omnivores
 - b. Herbivores
 - c. Carnivores

12. Is Food Chain important? Why? Or Why not? Write your answer below.

Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.

Thank you for your help. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix B-English Reading Comprehension's Ordered Scoring Scheme

Reading Comprehension Item		Least correct	Partly correct or partly incorrect	Most correct
		Score 1	Score 2	Score 3
Item difficulty				
Need for Energy				
1	What will happen if there is no food?	a	b	c
2	A food chain is	a	c	b
3	Where do plants get energy from?	c	a	b
Parts of the Food Chain				
4	What are primary consumers?	c	b	a
5	Primary consumers are also called	a	b	c
6	What is an example of herbivores?	c	a	b
Secondary Consumers in the Food Chain				
7	What are carnivores?	a	c	b
8	Decomposers are	c	a	b
9	Which statement is <u>correct</u> ?	a	b	c
Humans in the Food Chain				
10	What do humans eat?	a	b	c
11	What do we call humans?	b	c	a
12	Is Food Chain important? Why?			

Appendix C-English Writing Test

Name:

Name of School:

Name of Class:

.....

.....

.....

WRITING

TOPIC 1. : *My Idul Fitri Holiday*

Direction: 1. Put your name, name of school, and name of class on the top right of this page.

2. Write some paragraphs (the more paragraphs the better) on '*My Idul Fitri Holiday*'

Time: 40 minutes

My Idul Fitri Holiday

5

10

15

TOPIC 2. : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ‘ My Family’.

Time: 40 minutes

My Family

5

10

15

20

Appendix D-English Writing Scoring Rubric

		TEXT WRITING	EXCELLENT (Score 4)	GOOD (Score 3)	ADEQUATE (Score 2)	POOR (Score 1)
Paragraph(s) Organisation						
Easy	1	Main Topic	Main topic is clearly stated in a complete opening sentence with correct sentence structure.	Main topic is stated in an opening sentence but sentence structure is not correct AND/OR lacking detail.	Main topic is not clearly stated but sentence structure is correct.	Main topic is not clearly stated. Sentence structure is not correct.
Hard	2	Concluding Sentence	Concluding sentence restates topic with correct sentence structure.	Concluding sentence restates topic but sentence structure is not correct AND/OR lacking detail.	Concluding sentence is not clearly stated but sentence structure is correct.	Concluding sentence is not clearly stated. Sentence structure is not correct.
Harder	3	Supporting Sentences	Text has at least 3 detailed supporting sentences AND all sentences are on topic.	Text has at least 3 supporting sentences *Only 2 sentences are on topic. AND/OR *Sentences lack details	Text has at least 3 supporting sentences *Only 1 sentence is on topic. *Sentences lack details.	Text has supporting details BUT none are on topic.
Text Conventions						
Easy	4	Spelling	There are 0-2 spelling errors.	There are 3-4 spelling errors.	There are 5-6 spelling errors.	There are more than 6 spelling errors.
Hard	5	Punctuation/ Capitalisation	Text has 0-2 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 3-4 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 5-6 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has more than 6 errors in punctuation/ capitalisation, and noun-verb agreement.
Harder	6	Grammar	Text has 0-2 errors in noun-verb agreement.	Text has 3-4 errors in noun-verb agreement.	Text has 5-6 errors in noun-verb agreement.	Text has more than 6 errors in noun-verb agreement.
Text Quality						
Easy	7	Readability	Text is neat and readable with 0-2 marked out words or other corrections.	Text is neat and readable with 3-4 marked out words or other corrections.	Text is not neat and readable with 5-6 marked out words or other corrections.	Text is not neat and unreadable with numbers of marked out words or other corrections.
Hard	8	Style (Sentence fluency, e.g. varied length,	Text shows sentence fluency.	Text shows reasonable sentence fluency.	Text shows minimal sentence fluency.	Text lacking in sentence fluency.

		good flow rhythm, and varied structure)				
Harder	9	Text Enjoyability	Text is enjoyable to read	Text is quite enjoyable to read	Text is satisfactory to read	Text is not enjoyable to read

Appendix E-Attitude/Behaviour Questionnaire

Name	
------	--



QUESTIONNAIRE:

Students' attitudes and behaviour questionnaire in relation to learning English.

Please put your name in the box on right top of this page.

Please rate the 18 statements according to the following response format and place and check (√) the column corresponding to your attitude (prior to studying), your behaviour (during study) on appropriate line opposite each statement:

- | | |
|----------------------|---------------------------|
| Most or all the time | put check (√) on column 3 |
| Some of the time | put check (√) on column 2 |
| Never or rarely | put check (√) on column 1 |

Instruction: Check (√) in the one column of the 3 scales that you feel most describes your feeling. Any response in the questionnaire will not affect your score in the course.

Subheading: Tasks for Listening Example

Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English			(√)		(√)	

Check (√) on the questionnaire item, if you believe, prior to learning that *you most or all the time like to pay attention to someone speaking English*, check (√) on column 3; if *you some of the time achieve this*, check (√) on column 2.

Now, answer the questions on the next page.

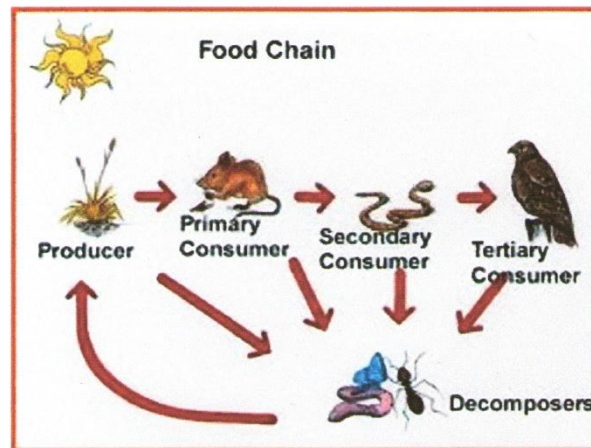
Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English						
3-4	I ask others to speak slowly or repeat words in English						
5-6	I listen to English songs						
	Tasks for Speaking						
7-8	I say new words several times in English						
9-10	I practice English with other students						
11-12	I start conversation in English with my friends						
	Tasks for Reading						
13-14	I guess the meaning of the English words in the text						
15-16	I read carefully words in English						
17-18	I can read English at home on my own						
	Tasks for Writing						
19-20	I look for similar words in my own language						
21-22	I divide English words into parts that I understand						
23-24	I write feelings in a diary in English						
	Student/student relationships						
25-26	I learn more when I study English in groups						
27-28	I can understand English better when I do activities with friends						
29-30	I like to lead my friends in doing English activities						
	Student/teacher relationships						
31-32	I learn a lot from my English teacher						
33-34	I like the way my teacher teaches English Reading Comprehension						
35-36	I like the way my teacher teaches English writing.						
	Common Views						
37-38	I like English because we use it in the classroom						
39-40	I like English because it helps me in higher study						
41-42	I like English because it helps me go abroad						

Is there anything else you would like to add about your experiences in learning English?

Thank you for your help.
It is appreciated. Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix F-Original English Reading Comprehension Text

FOOD CHAINS



Plants are important to our planet because they use photosynthesis to provide animal life with the energy they need to grow and to carry out life processes. When plants are eaten, this energy is transferred to another species. We call this a food chain.

There are three basic levels of a food chain.

Level 1—Plants are called producers because they have the ability to convert the Sun’s energy into chemicals that can be used to power life processes. Plants capture about 0.01 percent of the energy of the Sun’s rays every day and are the basis of the whole food chain.

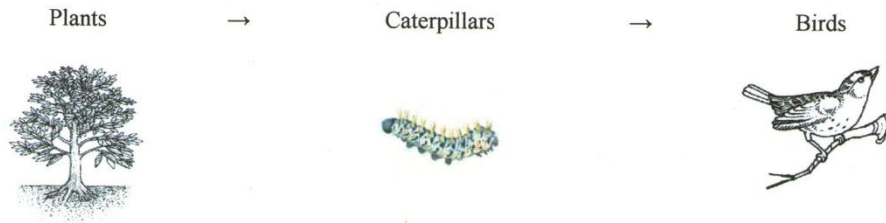
Level 2—Animals on the second level of a food chain are called primary consumers. We call them herbivores because they eat only plants for their energy. For example, deer eat shrub leaves, rabbits eat carrots and worms eat left litter. Primary consumers need to eat a lot of plants to meet their daily energy requirement. They eat approximately 10 per cent of the available plants and gain about 0.001 per cent of the Sun’s original energy.

Level 3—Animals on the third level are called secondary consumers. They eat primary consumers for their energy, for example, lions eat deer, foxes eat rabbits and birds eat worms. Secondary consumers are good predators, but there are fewer of them because their energy sources are limited. Secondary consumers usually eat meat and we call them carnivores (if secondary consumers eat mainly insects, we call them insectivores). Secondary consumers gain about 0.0001 per cent of the Sun’s original energy.

Further levels of the food chain are possible, but the longer a food chain is, the less energy left for the species at the top of the chain.

Representing food chains

We can represent food chains using arrows like this:



The arrow represents what species eat what.

That is:

Plants (eaten by) Caterpillar (eaten by) Birds

HUMANS AS CARNIVORES

Humans are at the top of their food chain because other animals do not naturally eat them. Humans are called omnivores because they can eat a variety of foods, such as plants and animals. Humans are not the largest of all species, but they have superior brains, they often eat animals that are larger than themselves. This is because they have used their intelligence to develop farming and hunting techniques.

[Source Adopted from: Walker, D (2006). *Adaptation and survival: Science Essential, Biology*. London: Evans]

DIRECTIONS:

1. Put **your name** on the **answer sheet** provided.
2. Read the passage carefully and cross (X) a, b, or c that suits your answer best.

QUESTIONS:

Plants in the Food Chain

1. To what does 'they' in line 1 refer?
 - a. Plants
 - b. Animals
 - c. Humans
2. A food chain is
 - a. a transfer of food from a species to another
 - b. a transfer of energy from a species to another
 - c. a transfer of chain from a species to another

Page | 3

Appendix G-Original Attitude/Behaviour Test

Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English						
3-4	I ask others to speak slowly or repeat words in English						
5-6	I listen to English songs						
	Tasks for Speaking						
7-8	I say new words several times in English						
9-10	I practice English with other students						
11-12	I start conversation in English with my friends						
	Tasks for Reading						
13-14	I guess the meaning of the English words in the text						
15-16	I read carefully words in English						
17-18	I can read English at home on my own						
	Tasks for Writing						
19-20	I look for similar words in my own language						
21-22	I divide English words into parts that I understand						
23-24	I write feelings in a diary in English						
	Student/student relationships						
25-26	I learn more when I study English in groups						
27-28	I can understand English better when I do activities with friends						
29-30	I like to lead my friends in doing English activities						
	Student/teacher relationships						
31-32	I learn a lot from my English teacher						
33-34	I like the way my teacher teaches English Reading Comprehension						
35-36	I like the way my teacher teaches English writing.						

Is there anything else you would like to add about your experiences in learning English?

Thank you for your help.

It is appreciated. Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix H-Request Letter to Head of Provincial Education of Aceh



September 27, 2010

Head
Dinas Pendidikan Aceh
in
Banda Aceh

Dear Sir,

With regard to activities for a PhD data collection which will be conducted by my PhD student, as the following details:

Name	:	Khairiah Syahabuddin
Student Number	:	10161350
Programme	:	Doctor of Philosophy (PhD)
Faculty	:	Education and Arts
University	:	Edith Cowan University, Mount Lawley Campus, Perth, Western Australia.

I would like to request that you can give permission to Khairiah Syahabuddin to collect her research data at Junior High School (Sekolah Menengah Pertama) in Banda Aceh for 1 (one) year, which meets a research procedure of Edith Cowan University, Australia.

Thank you. Your help and participation are appreciated.

A handwritten signature in cursive script, reading 'Russell Waugh', is positioned below the text of the letter.

Professor Russell Waugh

Principal Supervisor,
Edith Cowan University,
Mount Lawley Campus, 6050
Perth, Western Australia
Email: r.waugh@ecu.edu.au

Appendix I-Consent Letter to Headmaster



Dear Headmaster,

I am conducting a research project on student achievement in English Reading Comprehension; English writing; and Attitude and Behaviour.

The purpose of the project is to obtain information on your student achievement in English Reading Comprehension, English writing, and Attitude and Behaviour.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation of your students is voluntary and they can pull out at any time without prejudice. Your student participation has nothing to do with any formal or informal assessment in their school subjects.

Your student' name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of the students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

August 2010

I consent my students to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix J-Consent Letter to English Teacher



Dear Classroom Teacher,

I am conducting a research project on student achievement in English Reading Comprehension; English writing; and Attitude and Behaviour.

The purpose of the project is to obtain information on your student achievement in English Reading Comprehension, English writing, and Attitude and Behaviour.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation of your students is voluntary and they can pull out at any time without prejudice. Your student participation has nothing to do with any formal or informal assessment in their school subjects.

Your student' name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of the students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

August 2010

I consent my students to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix K-Consent Letter to student (English Reading Comprehension)



Dear Student,

I am conducting a research project on student achievement in English reading comprehension.

You are asked to complete the attached reading comprehension test. It contains 12 questions in multiple choices within 40 minutes.

The purpose of the test is to obtain information on your achievement in reading comprehension.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the test is given on the conditions mentioned above. Please return the completed test to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University (ECU), School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix L-Consent Letter to Student (English Writing)



Dear Student,

I am conducting a research project on student achievement in English writing.

You are asked to write some paragraphs (the more paragraphs is better) on two assigned topics. Each topic is allocated 35 minutes.

The purpose of the test is to obtain information on your achievement in writing.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the test is given on the conditions mentioned above. Please return the written paragraphs to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix M-Consent Letter to Student (Attitude/Behaviour Questionnaire)



Dear Student,

I am conducting a research project on student attitude and behaviour in learning English.

You are asked to complete the attached questionnaire. It contains 18 statements covering six aspects of your attitudes and behaviour in relation to learning English. These are tasks for listening, tasks for speaking, tasks for reading, tasks for writing, student/student relationships, and student/teacher relationships.

The purpose of the test is to obtain information on your attitude and behaviour in learning English.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@student.ecu.edu.au or khairiah_syahabuddin@yahoo.com

If you wish to speak with an independent person about the conduct of the project, please contact Ms Kim Gifkins, Research Ethics Officer, Edith Cowan University, on +61863042170, or research.ethics@ecu.edu.au

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the questionnaire is given on the conditions mentioned above. Please return the completed questionnaire to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University (ECU), School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signature:

Name: _____

Date: _____

School name: _____

Appendix N-Student's Sample Answer of Reading Comprehension

ANSWER SHEET for Reading Comprehension 'FOOD CHAINS'

Name: _____

Name of School: _____

Name of Class: _____

Please answer here:

Read the passage carefully and cross (X) a, b, or c that suits your answer best.

Needs for Energy

1. a. b. c.
2. a. b. c.
3. a. b. c.

Sunlight and Producers

4. a. b. c.
5. a. b. c.
6. a. b. c.

Consumers and Decomposers

7. a. b. c.
8. a. b. c.
9. a. b. c.

Humans in the Food Chains

10. a. b. c.
11. a. b. c.

12. Is Food Chain important? Why? Or Why not? Write your answer below.

Yes, Because ~~living~~ living things need food to give their energy.

Plants get energy from sunlight. The plants and animals in a food chain depend on each other to live. So, every part of the chain is important.

Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.

Yes, I like ^{english} reading comprehension.

Because:

1. I like english language.

2. English Reading Comprehension make me know about ~~comper~~ comprehension.

3. make me to be know much information

Thank you for your help. It is appreciated.
Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

**Appendix O-Student's Sample Answer English Writing
(Bilingual Student)**

Name: [Redacted]
Name of School: [Redacted]
Name of Class: [Redacted]

WRITING

TOPIC 1. : My Idul Fitri Holiday

Direction: 1. Put your name, name of school, and name of class on the top right of this page.
2. Write some paragraphs (the more paragraphs the better) on 'My Idul Fitri Holiday'

Time: 40 minutes

My Idul Fitri Holiday

5 Last year, I spent my Idul Fitri holiday in my village. I'm very happy in there. Because, I can play with my cousins and my friends. We're playing mud on the ricefield and we also made homework together. I have two cousins, they are very nice to me. We also celebrate my brother's birth on my village, and visited my grandfather's kuburan.

On first week of holiday, we was visit to my aunt house in Lhokseumawe. In there, we meet our family. With this time, I also to shop and buy many things in Suzuya store. Lhokseumawe is nice and good city. We can see fire on Pt. Arun, the fire is so beautiful.

10 Second week, we're back to Banda Aceh. Many memories to remember the holiday. Before that, I don't want to back to Banda Aceh. Because, I fell very happy in my village. I love my village.

15

TOPIC 2 : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ' My Family'.

Time: 40 minutes

My Family

- Hi... I want to tell you about my family. My name is [REDACTED].
I have a nicest family. My father's name is Asnawi Ismail. His job
is journalist on Harian Serambi Indonesia, and he is a lecture on
IAIN Ar-Raniry University. My father's favourite food is canai, same
5 to me. Because, canai is delicious food. Canai's from India. But,
we can buy that in "Canai Mamak Cafe". My father was born on
october, 10th, 1965.
- Now, I want to talk about my mother. My mother's name is Nur Azzah, sh.
She is a civil servant on KPU office. She has white skin and slim body.
10 Her hair is black and wavy. My mother's very like to make me many
food. like pizza, crocet, spaghetti, and many food's delicious. My mother
wear glasses, same to my father. Every day, my mother make me delicious
and healthy food.
- And one more, I have a little brother, his name is Hafiz Maulana.
15 He is 10 years old, he was born on January, 27th, 2001. He is
a smart boy, His school is elementary school number 67 Percontohan,
Banda Aceh. now, he is grade four. My little brother has pointed nose and
white skin. He's very handsome and nicest boy. He's very like otomotif,
like my uncle. In his bedroom, many toys of otomotif his collection.
- 20 Ok, guys, That's about my family. My family live at complex perumahan
Puklat Uke Kareng. If you want to Ulee Kareng, you can visit my house.
Thank you.

**Appendix P-Student's Sample Answer English Writing
(Monolingual Student)**

Name: _____
Name of School: _____
Name of Class: _____

WRITING

TOPIC 1. : My Idul Fitri Holiday

Direction: 1. Put your name, name of school, and name of class on the top right of this page.
2. Write some paragraphs (the more paragraphs the better) on 'My Idul Fitri Holiday'

Time: 40 minutes

My Idul Fitri Holiday

My Name is _____, I am a student ~~the~~ ~~STH~~ of junior high school three, I am My Address is Ateuk Munjeng No. 5. my Father job is PNS (teacher), My Mother job is PNS (teacher).
at the feast of Idul Fitri, our family went to the mosque
5 after returning from the mosque we went to the house the young Sister to stay in touch. on our second holiday to return home. we went ^{up} ~~by~~ the car after I got ~~to~~ there, we greet shaking hands with the ~~gran~~ grand mother, grand father. and other relatives.

10

15

TOPIC 2 : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ' My Family'.

Time: 40 minutes

My Family

My Name is [REDACTED], My Mother ~~is~~ my name is
***** , my father name is ***** I have 2 young
Sister young sister 1. my name is NISA AUFIA , young sister 2 my name
is FATIN NAFISA, I am a student of junior-high school three
5 I am my adress is Ateul- Muriheng No. 5. is father Job.
PNS (teacher), My Mother job is (teacher matematika).
I happy the family. I am have unke my name is YAYAU
I am HAVE Aunt my Name is Xanti. I am is
very always swimming the pool - I Hobby Reading ~~book~~ Story
10 Book Malin Kundang, Laskar Pelangi, EVIL
old I am ~~the~~ ~~twelve~~ , grand mother my name is NUSJANAH
My grand father Rusli M. ALI

15

20

Appendix Q-Student's Sample Answer Attitude/Behaviour Questionnaire (Bilingual Student)

Nama [REDACTED] 1251
 Nama Sekolah [REDACTED]
 Nama Kelas [REDACTED]

Pernyataan Kuesioner		Ini yang saya pikir seharusnya terjadi			Ini yang sebenarnya terjadi		
		1	2	3	1	2	3
Nomor Pernyataan							
Kegiatan kegiatan dalam Listening (mendengarkan)							
1-2	Saya mendengarkan dengan benar pada orang yang sedang berbicara Bahasa Inggris.			✓		✓	
3-4	Saya meminta orang lain untuk berbicara dengan lambat atau meminta untuk mengulangi kata kata ketika berbicara Bahasa Inggris.	✓				✓	
5-6	Saya mendengarkan lagu-lagu Bahasa Inggris.			✓			✓
Kegiatan-Kegiatan dalam Speaking (Berbicara)							
7-8	Saya mengucapkan kata kata baru Bahasa Inggris beberapa kali.			✓		✓	
9-10	Saya berlatih bicara Bahasa Inggris dengan siswa lain.			✓		✓	
11-12	Saya memulai percakapan Bahasa Inggris dengan teman-teman saya.			✓		✓	
Kegiatan-Kegiatan dalam Reading Comprehension (Membaca dan Memahami)							
13-14	Saya menebak arti dari kata kata Bahasa Inggris dalam teks.		✓			✓	
15-16	Saya membaca kata kata Bahasa Inggris dengan seksama.			✓		✓	
17-18	Saya dapat membaca Bahasa Inggris sendiri di rumah.			✓		✓	
Kegiatan-Kegiatan dalam Writing (Menulis)							
19-20	Saya mencari kata kata Bahasa Inggris yang mirip dengan kata-kata Bahasa Indonesia ketika menulis Bahasa Inggris.	✓				✓	
21-22	Saya menulis kalimat Bahasa Inggris yang pendek pendek.	✓				✓	
23-24	Saya menulis perasaan saya ke dalam diari dalam Bahasa Inggris.			✓		✓	
Interaksi Siswa dengan Siswa							
25-26	Saya belajar banyak tentang Bahasa Inggris dalam kelompok belajar.			✓			✓
27-28	Saya dapat mengerti Bahasa Inggris dengan lebih baik ketika saya melakukan aktivitas dengan teman-teman saya.			✓		✓	
29-30	Saya suka memimpin teman-teman melakukan aktifitas Bahasa Inggris.			✓		✓	
Interaksi Siswa dengan Guru							
31-32	Saya belajar banyak tentang Bahasa Inggris dari guru Bahasa Inggris saya.			✓			✓
33-34	Saya suka cara guru Bahasa Inggris saya mengajar Membaca dan Memahami (Reading Comprehension).			✓			✓
35-36	Saya suka cara guru Bahasa Inggris saya mengajar Menulis (Writing).			✓			✓
Pandangan Umum							
37-38	Saya suka Bahasa Inggris karena kami menggunakannya di kelas.			✓			✓
39-40	Saya suka Bahasa Inggris karena dapat membantu saya ke sekolah lanjutan.			✓			✓
41-42	Saya suka Bahasa Inggris karena dapat membantu saya pergi ke luar negeri.			✓			✓

Ada hal-hal lain yang ingin kamu tambahkan tentang pengalamanmu dalam belajar Bahasa Inggris?

Orang tua saya sering membantu dalam belajar bahasa Inggris, misalnya Ibu saya sering mengajak saya berbicara bahasa Inggris, dan Ayah saya sering membelikan saya buku cerita berbahasa Inggris. Teman-teman saya di Facebook juga ada beberapa yg

Terima Kasih. Bantuanmu sangat dihargai. berasal dari luar negeri, jadi Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL) saya ber-chatting dengan mereka menggunakan bahasa Inggris.

Appendix R-Student's Sample Answer Attitude/Behaviour Questionnaire (Monolingual Student)

Nama _____
 Nama Sekolah _____
 Nama Kelas _____

Pernyataan Kuesioner		Ini yang saya pikir seharusnya terjadi			Ini yang sebenarnya terjadi		
		1	2	3	1	2	3
Nomor Pernyataan							
Kegiatan kegiatan dalam Listening (mendengarkan)							
1-2	Saya mendengarkan dengan benar pada orang yang sedang berbicara Bahasa Inggris.		✓				✓
3-4	Saya meminta orang lain untuk berbicara dengan lambat atau meminta untuk mengulangi kata kata ketika berbicara Bahasa Inggris.		✓			✓	
5-6	Saya mendengarkan lagu-lagu Bahasa Inggris.			✓			✓
Kegiatan-Kegiatan dalam Speaking (Berbicara)							
7-8	Saya mengucapkan kata kata baru Bahasa Inggris beberapa kali.		✓			✓	
9-10	Saya berlatih bicara Bahasa Inggris dengan siswa lain.	✓				✓	
11-12	Saya memulai percakapan Bahasa Inggris dengan teman-teman saya.	✓			✓		
Kegiatan-Kegiatan dalam Reading Comprehension (Membaca dan Memahami)							
13-14	Saya menebak arti dari kata kata Bahasa Inggris dalam teks.		✓				✓
15-16	Saya membaca kata kata Bahasa Inggris dengan seksama.		✓			✓	
17-18	Saya dapat membaca Bahasa Inggris sendiri di rumah.			✓			✓
Kegiatan-Kegiatan dalam Writing (Menulis)							
19-20	Saya mencari kata kata Bahasa Inggris yang mirip dengan kata-kata Bahasa Indonesia ketika menulis Bahasa Inggris.	✓			✓		
21-22	Saya menulis kalimat Bahasa Inggris yang pendek pendek.		✓		✓		
23-24	Saya menulis perasaan saya ke dalam diari dalam Bahasa Inggris.		✓				✓
Interaksi Siswa dengan Siswa							
25-26	Saya belajar banyak tentang Bahasa Inggris dalam kelompok belajar.		✓			✓	
27-28	Saya dapat mengerti Bahasa Inggris dengan lebih baik ketika saya melakukan aktivitas dengan teman-teman saya.		✓		✓		
29-30	Saya suka memimpin teman-teman melakukan aktifitas Bahasa Inggris.			✓			✓
Interaksi Siswa dengan Guru							
31-32	Saya belajar banyak tentang Bahasa Inggris dari guru Bahasa Inggris saya.			✓			✓
33-34	Saya suka cara guru Bahasa Inggris saya mengajar Membaca dan Memahami (Reading Comprehension).		✓			✓	
35-36	Saya suka cara guru Bahasa Inggris saya mengajar Menulis (Writing).		✓			✓	
Pandangan Umum							
37-38	Saya suka Bahasa Inggris karena kami menggunakannya di kelas.		✓		✓		
39-40	Saya suka Bahasa Inggris karena dapat membantu saya ke sekolah lanjutan.			✓			✓
41-42	Saya suka Bahasa Inggris karena dapat membantu saya pergi ke luar negeri.			✓			✓

Ada hal-hal lain yang ingin kamu tambahkan tentang pengalamanmu dalam belajar Bahasa Inggris?

Untuk menambah wawasan saya, Untuk bisa membaca komik dan novel, mengerti arti dari lagu Inggris. Dan untuk membantu saya di masa depan.

Terima Kasih. Bantuanmu sangat dihargai.
 Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix S-Complete Occurrence of Student Written Comment

THEMES/CONCEPTS OF STUDENTS' WRITTEN COMMENTS

Themes/Concepts	Bilingual	Monolingual	Total
1. Tasks on Learning English as a Foreign Language			
Tasks for Listening			
I pay attention to someone speaking English	1	1	2
I ask others to speak slowly or repeat words in English	3	0	3
I like listening to English songs	25	19	44
Listening helps me to write	1	0	1
TOTAL			50
Tasks for Speaking			
I practice English with other students	9	1	10
It is important to be able to converse in English fluently	3	1	4
I like speaking in English	21	10	31
I like sing songs	3	1	4
TOTAL			49
Tasks for Reading			
I guess the meaning of the English words in the text	4	0	4
I like learning English through reading comprehension	16	8	24
I like reading books/e-books in English	5	0	5
TOTAL			33
Tasks for Writing			
I write feelings in a diary in English	1	1	2
I like to create a new sentence with a new vocabulary	1	0	1
I like to use dictionary for new words	1	1	2
I understand written text more than spoken	1	0	1
I like writing	2	2	4
TOTAL			10
2. Student-student relationships			
I learn more when I study English in groups	13	7	20
I ask friends to help me learning English	8	1	9
TOTAL			29
3. Student-teacher relationships			
I learn a lot from my English teacher	13	5	18
I like the way my teacher teaches English	57	41	98
I don't like the way my teacher teaches English	4	12	16
The way teacher teaches gives impact to students to love or hate English	9	3	12
TOTAL			144

4. Personal Views of Learning English as a Foreign Language			
Learning English is easy	15	5	20
Learning English is important	9	12	21
I like English because it is unique (it is pleased on ears but difficult to pronounce)	7	9	16
I like learning English	96	70	166
I like English because it is challenging	19	9	28
I like English because it is enjoyable/fun	74	50	124
I like English because number of good books are in English	1	0	1
I want to learn English with simple sentence	1	2	3
I am happy with my vocabulary	1	3	4
English is enjoyable if studied correctly	2	0	2
English is my favourite subject	1	0	1
Learning English needs time	2	1	3
I like making story in English	1	0	1
Other language helps understand English	1	1	2
I like to translate	0	3	3
I like storytelling	2	0	2
Dictionary helps me	7	3	10
TOTAL	408		
5. Common Views on Benefit of Learning English as a Foreign Language			
I like English because it helps me in higher study	76	59	135
I like English because it helps me go abroad/around the world	77	66	143
I like English because it helps me to study abroad	17	6	23
I like English because it helps me get scholarship for higher studies	1	0	1
I like English because it is an international language	46	41	87
I like English because it helps me in my daily activities	16	9	25
I like English as an additional language	3	7	10
I like English because it helps me to be enrolled in favourite schools	2	1	3
I like English because we use it in the classroom	6	0	6
I like English because it is a means to pursue my future dream jobs	48	27	75
TOTAL	508		
6. Students' Confidence and Achievement			
I am motivated to learn English	151	87	238
I take after-school English lesson to improve my English	18	7	25
I like English since I was a kid	4	0	4
I am lucky to understand English	1	0	1
I won English competition	11	7	18
TOTAL	286		

7. Learning English through Media			
I like watching TV programs in English for kids	1	0	1
I like watching TV programs/films in English	5	2	7
I like paying attention in found conversation in English movies	1	0	1
I like remembering sentences found in English movies	1	0	1
I like playing games/songs/story-book to improve my English	12	5	17
I like English because numbers of media/software/games/computers/video use English	7	2	9
I like learning English through English laboratory	1	0	1
TOTAL	37		
8. Family Support in Learning English as a Foreign Language			
My family (father, mother, uncles, aunts, siblings) encourage me to learn English well for future studies	37	6	43
I like to practice English with my older sister/brother	33	12	45
I like learning English to make my parents proud	0	7	7
TOTAL	95		
9. Obstacles in Learning English as a Foreign Language			
English is not easy	6	22	28
I do not understand vocab/meanings/grammar in English much	15	21	36
I do not like English because it is not easy to learn	31	41	72
I lack in English listening	2	3	5
I lack in English speaking	19	4	23
I lack in English writing	4	0	4
I lack in English reading comprehension	2	0	2
I do not like English, beside we do not need English in this country	4	0	4
I wish I knew English/good at it	3	5	8
I am not good in English	5	2	7
TOTAL	189		
10. Other Views			
The questionnaire gives me insights on my perceptions on English	2	0	2
The test given (by the researcher) is the first one considered well-prepared and presented	1	0	1
The test gave me lesson.	1	1	2
TOTAL	5		

1110 732 1846

LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW

HKI

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia

Penulis : Khairiah Syahabuddin

Pengusul : Khairiah Syahabuddin

Status Pengusul : Pengusul tunggal

Identitas HKI : a. No. Sertifikat : 02330
b. No. dan Tanggal Permohonan : EC00201701407, 22 Mei 2017
c. Penerbit : Kemenkumham RI
d. Tahun : 2017
e. Link Penerbit : <https://pdki-indonesia.dgip.go.id/detail/EC00201701407?type=copyright&keyword=Achievement%2C+attitude%2C+and+behavior+in+bilingual>

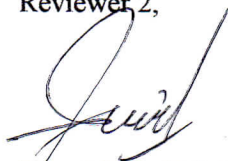
Kategori : Sertifikat HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
Sertifikat HKI, Nomor Pencatatan 02330	10	10
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

Banda Aceh, 21 Februari 2022
Reviewer 2,



Dr. phil, Saiful Akmal, M.A.
NIP. 198203012008011006

Lektor Kepala Bidang Pendidikan Bahasa Inggris pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

**LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW
HKI**

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
 Penulis : Khairiah Syahabuddin
 Pengusul : Khairiah Syahabuddin
 Status Pengusul : Pengusul tunggal
 Identitas HKI : a. Penerbit : Kemenkumham RI
 b. Tahun : 2017

Kategori : Konten HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
e. Kelengkapan unsur isi HKI (10%)	1	1
f. Ruang lingkup dan kedalaman pembahasan (30%)	3	3
g. Kecukupan dan kemutakhiran data/informasi dan metodologi (30%)	3	3
h. Kelengkapan unsur dan kualitas penerbit (30%)	3	3
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

- a. Kelengkapan unsur isi HKI:
sangat lengkap
-
- b. Ruang lingkup dan kedalaman pembahasan:
Lengkap dan mendalam
-
- c. Kecukupan dan kemutakhiran data/informasi dan metodologi:
Data/informasi dan metodologi mutakhir
-
- d. Kelengkapan unsur dan kualitas penerbit:
Lengkap dan berkualitas
-

Banda Aceh, 21 Februari 2022

Reviewer 2,

Dr. phil, Saiful Akmal, M.A.

NIP. 198203012008011006

Lektor Kepala Bidang Pendidikan Bahasa Inggris pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh



REPUBLIK INDONESIA
KEMENTERIAN HUKUM DAN HAK ASASI MANUSIA

SURAT PENCATATAN CIPTAAN

Menteri Hukum dan Hak Asasi Manusia Republik Indonesia, berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta yaitu Undang-Undang tentang perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra (tidak melindungi hak kekayaan intelektual lainnya), dengan ini menerangkan bahwa hal-hal tersebut di bawah ini telah tercatat dalam Daftar Umum Ciptaan:

- I. Nomor dan tanggal permohonan : EC00201701407, 22 Mei 2017
- II. Pencipta
Nama : **Khairiah Syahabuddin, MHSc.ESL., M.TESOL., Ph.D**
Alamat : Jl. Got-Got No. 2, Sektor Utara, Kopelma Darussalam, Banda Aceh, DI ACEH, 23111
Kewarganegaraan : Indonesia
- III. Pemegang Hak Cipta
Nama : **Khairiah Syahabuddin, MHSc.ESL., M.TESOL., Ph.D**
Alamat : Jl. Got-Got No. 2, Sektor Utara, Kopelma Darussalam, Banda Aceh, DI ACEH, 23111
Kewarganegaraan : Indonesia
- IV. Jenis Ciptaan : Karya Tulis
- V. Judul Ciptaan : **Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia**
- VI. Tanggal dan tempat diumumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indonesia : 12 Desember 2013, di Perth
- VII. Jangka waktu perlindungan : Berlaku selama hidup Pencipta dan terus berlangsung selama 70 (tujuh puluh) tahun setelah Pencipta meninggal dunia, terhitung mulai tanggal 1 Januari tahun berikutnya.
- VIII. Nomor pencatatan : 02330

Pencatatan Ciptaan atau produk Hak Terkait dalam Daftar Umum Ciptaan bukan merupakan pengesahan atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang dicatat. Menteri tidak bertanggung jawab atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang terdaftar. (Pasal 72 dan Penjelasan Pasal 72 Undang-undang Nomor 28 Tahun 2014 Tentang Hak Cipta)

a.n. MENTERI HUKUM DAN HAK ASASI MANUSIA
REPUBLIK INDONESIA
DIREKTUR JENDERAL KEKAYAAN INTELEKTUAL
u.b.
DIREKTUR HAK CIPTA DAN DESAIN INDUSTRI

Dr. Dra. Erni Widhyastari, Apt., M.Si.
NIP. 196003181991032001

LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW

HKI

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
Penulis : Khairiah Syahabuddin
Pengusul : Khairiah Syahabuddin
Status Pengusul : Pengusul tunggal
Identitas HKI : a. No. Sertifikat : 02330
b. No. dan Tanggal Permohonan : EC00201701407, 22 Mei 2017
c. Penerbit : Kemenkumham RI
d. Tahun : 2017
e. Link Penerbit : <https://pdki-indonesia.dgip.go.id/detail/EC00201701407?type=copyright&keyword=Achievement%2C+attitude%2C+and+behavior+in+bilingual>

Kategori : Sertifikat HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
Sertifikat HKI, Nomor Pencatatan 02330	10	10
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

Sangat bagus dan sesuai

Banda Aceh, 21 Februari 2022
Reviewer 1,


Prof. Dr. T. Zulfikar, M.Ed.

NIP. 197804302001121002

Guru Besar/Profesor Bidang English Language Research pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW

HKI

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
Penulis : Khairiah Syahabuddin
Pengusul : Khairiah Syahabuddin
Status Pengusul : Pengusul tunggal
Identitas HKI : a. No. Sertifikat : 02330
b. No. dan Tanggal Permohonan : EC00201701407, 22 Mei 2017
c. Penerbit : Kemenkumham RI
d. Tahun : 2017
e. Link Penerbit : <https://pdki-indonesia.dgip.go.id/detail/EC00201701407?type=copyright&keyword=Achievement%2C+attitude%2C+and+behavior+in+bilingual>

Kategori : Sertifikat HKI

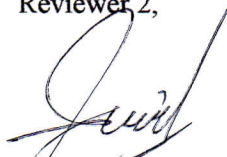
Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
Sertifikat HKI, Nomor Pencatatan 02330	10	10
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

Sudah sesuai

Banda Aceh, 21 Februari 2022
Reviewer 2,



Dr. phil, Saiful Akmal, M.A.

NIP. 198203012008011006

Lektor Kepala Bidang Pendidikan Bahasa Inggris pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

LINK DISERTASI DARI EDITH COWAN UNIVERSITY
(ECU) AUSTRALIA:

<https://ro.ecu.edu.au/theses/1083/>

2013

Student English achievement, attitude and behaviour in bilingual and monolingual schools in Aceh, Indonesia

Khairiah Syahabuddin
Edith Cowan University

Recommended Citation

Syahabuddin, K. (2013). *Student English achievement, attitude and behaviour in bilingual and monolingual schools in Aceh, Indonesia*. Retrieved from <https://ro.ecu.edu.au/theses/1083>

This Thesis is posted at Research Online.
<https://ro.ecu.edu.au/theses/1083>

Theses

Theses: Doctorates and Masters

Edith Cowan University

Year 2013

Student English achievement, attitude
and behaviour in bilingual and
monolingual schools in Aceh, Indonesia

Khairiah Syahabuddin
Edith Cowan University, ksyahabu@our.ecu.edu.au

This paper is posted at Research Online.
<http://ro.ecu.edu.au/theses/1083>

Edith Cowan University

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Student English Achievement,
Attitude and Behaviour
in Bilingual and Monolingual Schools
in Aceh, Indonesia

Khairiah Syahabuddin

S. Ag. (Bachelor of Arts in Islamic Studies), State Institute for Islamic Studies
(IAIN Ar-Raniry), Banda Aceh, Indonesia
M.H.Sc. in ESL, International Islamic University, Malaysia
M. TESOL, University of Southern Queensland, Australia

This dissertation is submitted for the degree of
Doctor of Philosophy

School of Education, Faculty of Education and Arts
Edith Cowan University, Perth, Western Australia

December 2013

Copyright Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

- (i) incorporate without acknowledgement material previously submitted for a degree or diploma in any institution of higher education;
- (ii) contain any material previously published or written by another person except where due reference is made in the text; or
- (iii) contain any defamatory material.

I also grant permission for the library of Edith Cowan University to make duplicate copies of my thesis as required.

Signed

Khairiah Syahabuddin

Date: December 11, 2013

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

Motto

من عرف لغة قوم سلم من مكرهم

In Bahasa Indonesia:

*“Barang siapa menguasai bahasa suatu kaum,
maka ia akan selamat dari tipu daya mereka.”*

*(Sebagai syarah [penjelasan singkat] dari Hadits Rasulullah [shallallaahu alaihi
wasallam] tentang kisah Zaid bin Tsabit agar mempelajari bahasa Suryani
[bahasa kaum Yahudi])*

In Acehnese:

*“Baranggasöe carông basa sabôh sabôh kawôm,
maka ureungnyan jeut seulamat dari tipee ngôn muslihat kawôm nyan.”*

*(Sebagöe syarah dari Hadist Rasulullah [shallallaahu alaihi wasallam] teuntang kisah
Zaid bin Tsabit nak geumurunöe bahsa Suryani [bahsa kawôm Yahudi])*

In English:

*“Those who master the language of a people,
will be safe from their conspiracy.”*

*(As a short explanation from Hadits of Rasulullah [peace upon him] on the experience
of Zaid bin Tsabit in order for him to learn the Suryani language [Jewish])*

Abstract

Following the tsunami in 2004, the education system in Banda Aceh, Indonesia, was reconstructed and revitalised, and part of this involved foreign intervention in setting up bilingual schools alongside state-run monolingual schools. The purpose of this study is threefold. The first is to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English essay writing, English reading comprehension, and attitude and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual and monolingual schools). The second is to investigate attitude and behaviour of students with regard to the learning of English as a foreign language, especially regarding student ability in English. The third is to explore students' beliefs and perceptions regarding their experiences of learning English as a foreign language.

A number of linear unidimensional scales were created for each of the three variables using Rasch Measurement with the 2010 RUMM computer program. The construct validity of the three variables was tested by designing the items in ordered patterns of item difficulty which were compared with their Rasch-measured item difficulties, as a Science-like test of the structure of the variables. An experimental research design (pretest/posttest, control/experimental group) was used with Rasch-created linear measures of three variables: (1) a researcher-designed English Essay Test; (2) a researcher-designed Reading Comprehension Test; and (3) a researcher-designed Attitude/Behaviour Test about Learning English. Seven hundred and eighty male and female first-year middle school students (aged 12-13 years old), consisting of 394 students from bilingual schools and 386 students from monolingual schools, selected from a number of schools with bilingual programs and monolingual programs, were the respondents for this study. After two months of lessons, the two groups were compared on each of the three measures using ANCOVA and ANOVA. Students' written comments were collected in regards to their experiences of learning English as a foreign language.

The findings showed that bilingual students outperformed monolingual students in tests of English Reading Comprehension, English Writing and Attitude/Behaviour for both pretests and posttests. Female students achieved better results than male students in English Reading Comprehension, English Writing, and Attitude/Behaviour tests, both for pretests and posttest.

Research Papers Based on this Dissertation

Syahabuddin, K. & Waugh, R. F. (2012a). Differential item functioning of English reading comprehension in a second language context when fit to the Rasch model is not ideal [Peer-reviewed]. Paper presented at the Fifth International Conference on Probabilistic Models for Measurement, held at the University of Western Australia, Perth, 23-25 January 2012.

Syahabuddin, K. & Waugh, R. F. (2012b). Students' ability in English reading comprehension by gender and bilingual/monolingual learning: An Acehese case (Banda Aceh). Paper presented at the Western Australia Institute for Education Research (WAIER), held at Notre Dame University, Fremantle, 11 August 2012.

Syahabuddin, K. & Waugh, R. F. (2012c). Students' Behaviour towards learning English as a foreign language by gender and bilingual/monolingual learning: An Acehese case (Banda Aceh). Paper presented at the Applied Linguistics Association of Australia (ALAA) National Conference, held at Curtin University, Perth, 12-14 November 2012.

<http://humanities.curtin.edu.au/schools/EDU/education/alaa-proceedings.cfm>

Acknowledgement



In the Name of Allāh, the Most Gracious, the Most Merciful

Three years ago, at the beginning of 2010, when starting this research journey, I was picturing myself as a person trying to reach the peak of a mountain with the help of a thin rope. Not easy. Being a full-time student while raising and nurturing two little children, aged 8 and 2.5 then, is not easy. The completion of this dissertation, therefore, is possible because of the abundant help and support I received. I am so thankful.

The most and foremost thanks, *Alhamdulillahirabbil 'Ālamēn*, goes to my God, Allah *Subhānahuwa Ta'ālā*, the Lord of the Universe. It is because of His Love I managed to finalise this dissertation. Grand blessings and peace are upon Prophet Muhammad (peace upon him) who has brought mankind from the darkness to the lightness of religion and knowledge. I am one of many who is given an opportunity to enrich my religion and knowledge.

My greatest appreciation goes to Professor Russell Waugh, my principal supervisor, who accepted me as one of his postgraduate students. It is he who introduced me to Rasch Measurement, a powerful research breakthrough, which I did not know before. He taught and nurtured me from the very beginning until this dissertation was completed. That involves thinking, reading, discussions, writing drafts, revising and re-revising. He makes sure I am 'on track', with, "*Khairiah, when are you going to give a new chapter to Dr Russell?*" or "*Well, Khairiah, is this the week you have a new chapter ready for me?*" When drafts are ready for him to read, it is only one or two days waiting, for the drafts return quickly to me. A lightning supervisor he is. I am deeply grateful. My appreciation also goes to my co-supervisor, Dr Anne Thwaite, who helped me with interesting discussions and thorough corrections.

My great thanks are to ECU, its School of Education, Graduate Research Support, and SOAR. My sincere thanks to the ECU Research Writing Consultant, Dr Jo McFarlane, and to my ECU roommates at Mount Lawley Postgraduate Room 16.238: Sulfasyah Syaharuddin (Sulfa), Omolola Aina Ladele (Lola), and Steven Chia (Steve) who share tips and tricks regarding writing up a dissertation.

I am also blessed by having friends from Indonesia and my hometown, Aceh, who helped cheer me up through BBQ gatherings, park picnics, Moslem celebrations, chit-chats and others who made me feel that I was not alone. They were beside me to offer help.

I also thank my home university, Ar-Raniry State Islamic University, Banda Aceh who released me on my study leave to pursue this PhD. Also my gratitude goes to 18 State Middle Schools in Banda Aceh who participated in this study: their headmasters, English teachers, and students.

My deepest gratitude goes to the Australia Endeavour Scholarship which supported my study for four years.

My warmest thanks go to my family in Banda Aceh, the family of *Bapak* Syahabuddin Mahyiddin: My mother *Mamak* Cut Nurhayati, H. Ahmad; my older brothers: *Bang* Rusydi Syahabuddin, *Bang* Mursyid Syahabuddin, *Bang* Armia Syahabuddin, and my only and younger sister, *Adek* Fadhilah Syahabuddin. Thanks also are to my sisters and brother-in-law, and cousins. My sincere thanks for them all who have always prayed and supported me with anything that one can imagine.

The last and the highest thanks goes to my family: My husband, *Kakanda* Sa'i Baharuddin; my daughter, Farhanah Arifah Sa'i; and my son, Tijan al-Darary Sa'i, who allowed me to take this study in the first place and to experience life in Perth; who were with me in my highs and lows; and who agreed to share some of their weekends, public holidays and school holidays with my work on this dissertation.

I am deeply thankful.

May we all be blessed.

Dedication

This dissertation is dedicated to my sunshine

daughter, **Farhanah Arifah Sa'i**,
and son, **Tijan al-Darary Sa'i**,
who cherish their life by being bilinguals

And to my husband, **Kakanda Sa'i Baharuddin**,
who is starting to be a bilingual

Also, to my late father, **Bapak Syahabuddin Mahyiddin**,
and my mother, **Mamak Cut Nurhayati H. Ahmad**,
who encouraged me to be a bilingual

Table of Contents

Front Page	i
Copyright Declaration.....	iii
Use of Thesis.....	v
Motto.....	vii
Abstract	ix
Research Papers Based on this Dissertation	xi
Acknowledgement	xiii
Dedication	xv
Table of Contents	xvi
List of Tables	xxvi
List of Figures	xxx
List of Appendices	xxxiv
CHAPTER ONE	1
INTRODUCTION.....	1
Background	1
Aceh Province	1
Tsunami Aftermath	2
Education in Aceh Province.....	4
Bilingual and Monolingual Education	10
Rationale	12
Statement of Purpose	13
Research Questions	13

Significance.....	14
Limitations	15
Definition of Terms.....	17
Structure of the Thesis	23
CHAPTER TWO	25
BILINGUALISM and BILINGUAL EDUCATION	25
The Nature of Bilingualism.....	25
What does it mean to be a bilingual?.....	25
What are the Benefits of Being a Bilingual?	31
Is there any Disadvantage to being Bilingual?	34
The Nature of Bilingual Education	36
What is Bilingual Education?	36
What are the Existing Bilingual Education Models and Practices?	37
Brief History of Bilingual Education Worldwide.....	45
The Nature of Bilingual Education in Indonesia.....	50
What is the Position of Bahasa Indonesia?.....	50
Why is English in the Indonesian Curriculum?.....	50
Why Do They Have Bilingual Education in Indonesia?	51
Bilingual Education Program Ban in Indonesia (2013).....	52
What is the future of Bilingual Education in Indonesia?.....	53
A Review of Bilingualism and English as a Second or Foreign Language	53
Review of Second Language Acquisition.....	54
Findings on Experimental and Control Studies for Reading and Writing, and Attitude and Behaviour	58
Findings Using Rasch Measures and Control and Experimental Groups.....	63

Gaps in Indonesian Research: Bilingualism <i>versus</i> Monolingualism.....	64
CHAPTER THREE	67
CONCEPTUAL FRAMEWORK FOR BILINGUAL/MONOLINGUAL TEACHING, MEASUREMENT, AND THE VARIABLES	67
Conceptual Framework for Bilingual and Monolingual Teaching in Banda Aceh	67
Measurement	71
True Score Theory Measurement	71
Rasch Measurement.....	72
Variables	78
Measuring English Reading Comprehension	78
Measuring English Writing Achievement	80
Questionnaire on Attitude and Behaviour with regard to Learning English	83
CHAPTER FOUR	85
RESEARCH DESIGN	85
Strategy and Design	85
Mixed-Method Design	86
Samples	89
Instruments for English Tests	92
Questionnaire on Student Attitude/Behaviour	93
Instruments for students' written comments	94
Administration of the Instruments	94
Procedure	94
Pilot Studies.....	95
Ethical Considerations and Study Approval	100

Ethics Approvals.....	100
Study Approval.....	102
Control of Extraneous Variables in the Quasi-Experiments.....	102
Test Data Collection, and Data Entry.....	104
Test Data Collection.....	104
Data Entry.....	105
Data Analysis.....	105
Analysis of Reading Comprehension and Essay Writing Test data.....	105
Analysis of Attitude and Behaviour Questionnaire Data.....	106
Analysis of Students' Written Comments.....	107
CHAPTER FIVE.....	109
DATA ANALYSIS (PART 1).....	109
RASCH MEASUREMENT OF READING COMPREHENSION.....	109
Output Analysis.....	110
Standardised Fit Residuals.....	110
Principal component analysis of residuals.....	111
Item-trait interaction (dimensionality).....	111
Person Separation Index.....	112
Individual Item Fit.....	113
Threshold Values.....	114
Scoring Category Curve.....	114
Item Characteristic Curves.....	116
Person-Item Threshold Distribution (Targeting).....	118
Differential Item Functioning (DIF) by Gender.....	119
Differential Item Functioning (DIF) by Type of Language Instruction.....	121

Targeting by Gender and Type of Language Instruction	123
Scale of Item Difficulties.....	124
Discussion and Summary.....	125
CHAPTER SIX	129
DATA ANALYSIS (PART 2).....	129
EXPERIMENTAL RESULTS FOR READING COMPREHENSION	129
Results on Mixed Design ANOVA (Split-Plot Design ANOVA).....	129
Results on Two-Way ANCOVA.....	133
Summary	135
CHAPTER SEVEN.....	139
DATA ANALYSIS (PART 3).....	139
RASCH MEASUREMENT OF ENGLISH WRITING	139
Initial Analysis (Posttest).....	139
Output from the Final Rasch Analysis (Postest).....	144
Standardised Fit Residuals.....	144
Dimensionality and Item-Trait Interaction	145
Reliability	146
Item Statistics	146
Scoring Category Curve	147
Item Characteristic Curves	148
Person-Item Threshold Distribution (Targeting).....	150
DIF by Gender	151
DIF by Type of Language Instruction	154
Targeting by Gender and Type of Language Instruction	154

Scale of Item Difficulties.....	156
Pretest.....	156
Discussion and Summary.....	157
CHAPTER EIGHT	159
DATA ANALYSIS (PART 4).....	159
EXPERIMENTAL RESULTS FOR ENGLISH WRITING	159
Results for Mixed Design ANOVA (Split-Plot Design ANOVA)	160
Results for Two-Way ANCOVA.....	163
Summary	165
CHAPTER NINE	169
DATA ANALYSIS (PART 5).....	169
RASCH MEASUREMENT OF ATTITUDE/BEHAVIOUR	169
Initial Analysis	169
Output from Final Analysis.....	170
Standardised Fit Residuals.....	170
Dimensionality and Item-Trait Interaction	170
Person Separation Index	171
Individual Item Fit	171
Threshold Values	172
Scoring Category Curve	173
Item Characteristic Curves	174
Person-Item Threshold Distribution (Targeting).....	175
DIF by Gender	177
DIF by Type of Language Instruction	179

Targeting by Gender and Type of Language Instruction	181
Scale of Item Difficulties.....	183
Discussion and Summary.....	183
CHAPTER TEN.....	185
DATA ANALYSIS (PART 6).....	185
EXPERIMENTAL RESULTS FOR BEHAVIOUR.....	185
Results for Mixed Design ANOVA (Split-Plot Design ANOVA)	185
Results for Two-Way ANCOVA.....	189
Summary	191
CHAPTER ELEVEN.....	193
DATA ANALYSIS (PART 7).....	193
STUDENTS' WRITTEN COMMENTS.....	193
Theme 1: Tasks on Learning English as a foreign language	194
Tasks for Listening	194
Tasks for Speaking	195
Tasks for Reading.....	195
Tasks for Writing.....	195
Theme 2: Student-Student Relationships.....	196
Theme 3: Student-Teacher Relationships	197
Like the way English is taught	197
Dislike the way English is taught	198
Theme 4: Personal Views on the Benefits of Learning English	198
Theme 5: Common Views on the Benefits of Learning English	201
Theme 6: Confidence and Achievement in Learning English	203

Theme 7: Learning English Through the Media	205
Theme 8: Family Support in Learning English.....	206
Theme 9: Obstacles in Learning English	208
Theme 10: Other Views	210
Summary	211
CHAPTER TWELVE.....	213
DISCUSSION OF RESULTS AND IMPLICATIONS.....	213
Summary of the Present Study	214
Answering the Research Questions.....	215
Summary of the Findings	220
A. English Reading Comprehension	220
B. English Writing	221
C. Behaviour with regard to Learning English as a Foreign Language.....	221
Discussion about Bilingual Education	222
Implications for Policy and Practice	226
Implications for Policy: Indonesian Ministry of Education/Policy Makers	227
Implications for Practice.....	227
Implication for Classroom Practices.....	230
Implications for Parents	231
Implications for Further Research	231
Limitations of the Study.....	233
Managing Time and Extraneous issues	233
Managing Teacher Perceptions (through Interviews).....	235
Managing Classroom Observations.....	235
Measures were not Ideal and Need Improvement	235

REFERENCES	237
APPENDICES	251
Appendix A-English Reading Comprehension.....	251
Appendix B-English Reading Comprehension’s Ordered Scoring Scheme.....	255
Appendix C-English Writing Test	256
Appendix D-English Writing Scoring Rubric.....	258
Appendix E-Attitude/Behaviour Questionnaire.....	260
Appendix F-Original English Reading Comprehension Text.....	262
Appendix G-Original Attitude/Behaviour Test.....	264
Appendix H-Request Letter to Head of Provincial Education of Aceh.....	265
Appendix I-Consent Letter to Headmaster	266
Appendix J-Consent Letter to English Teacher	267
Appendix K-Consent Letter to Student (English Reading Comprehension)	268
Appendix L-Consent Letter to Student (English Writing).....	269
Appendix M-Consent Letter to Student (Attitude/Behaviour Questionnaire)	270
Appendix N-Student’s Sample Answer of Reading Comprehension	271
Appendix O-Student’s Sample Answer English Writing (Bilingual Student).....	272
Appendix P-Student’s Sample Answer English Writing (Monolingual Student).....	274
Appendix Q-Student’s Sample Answer Attitude/Behaviour Questionnaire (Bilingual Student).....	276
Appendix R-Student’s Sample Answer Attitude/Behaviour Questionnaire (Monolingual Student).....	277
Appendix S-Complete Occurrence of Student Written Comment.....	278

List of Tables

Table 1.1 <i>Formal Education of Indonesia</i>	6
Table 2.1 <i>Typology of Bilingualism Classical</i>	29
Table 2.2 <i>Type of Bilingual Education</i>	44
Table 2.3 <i>Brief History of Bilingual Education in the USA</i>	46
Table 2.4 <i>Difference Among the Four Perspectives</i>	54
Table 3.1 <i>Item Difficulty Scoring Scheme for Reading Comprehension Test</i>	79
Table 3.2 <i>English Writing Scoring Rubric</i>	82
Table 3.3 <i>Attitude and Behaviour Questionnaire</i>	84
Table 4.1 <i>Student Numbers by Control/Experimental Groups and by Gender in Control/Experimental Groups</i>	91
Table 5.1 <i>Overall Fit Statistics Reading Comprehension Measure (N=780, I=12)</i>	111
Table 5.2 <i>Item-Trait Interaction for Reading Comprehension Scale</i>	112
Table 5.3 <i>Locations, Standard Errors, Residuals and Chi-Squares for Reading Comprehension Items</i>	113
Table 5.4 <i>Item Thresholds-uncentralised (Item=12, Number=780) for Reading Comprehension Measure</i>	114

Table 5.5	
<i>Order of Difficulty of Items on the Linear Scale</i>	124
Table 6.1	
<i>Pretest and Posttest Scores for the Bilinguals and Monolinguals</i>	132
Table 6.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	132
Table 6.3	
<i>Interaction Effect and Main Effect of Reading Comprehension between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	134
Table 7.1	
<i>Writing Marking Example for topic 1 'My Idul Fitri Holiday'</i>	142
Table 7.2	
<i>Initial Marking of English Writing Essay as Items (Scored in terms of 18 Items for Compatibility with RUMM2030)</i>	143
Table 7.3	
<i>Final Marking of English Writing Essay as Items (Scored in terms of 12 Items for Compatibility with RUMM2030)</i>	144
Table 7.4	
<i>Overall Fit Statistics for the Writing Test (N=780, I=12)</i>	145
Table 7.5	
<i>Item-Trait Interaction for the Writing Scale</i>	146
Table 7.6	
<i>Locations, Standard Errors, Residuals and Chi-Squares for Writing Items</i>	171
Table 7.7	
<i>Order of Difficulty of Items on the Linear Scale</i>	156
Table 8.1	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	161
Table 8.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	162
Table 8.3	
<i>Interaction Effect and Main Effect of Writing between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	163

Table 9.1	
<i>Overall Fit Statistics for the Behaviour Measure (N=779, I=11)</i>	170
Table 9.2	
<i>Item-Trait Interaction for Behaviour Scale</i>	171
Table 9.3	
<i>Locations, Standard Errors, Residuals and Chi-Squares for Behaviour Items</i>	172
Table 9.4	
<i>Item Thresholds Uncentralised (Item=11, Number=779) for Behaviour Measure</i>	173
Table 9.5	
<i>Order of Difficulty of Items on the Linear Scale</i>	183
Table 10.1	
<i>Pretest and Posttest Scores for the Bilinguals and Monolinguals</i>	187
Table 10.2	
<i>Pretest and Posttest Scores for the Female and Male Students</i>	188
Table 10.3	
<i>Interaction Effect and Main Effect of Behaviour between the Bilingually-taught Group and the Monolingually-taught Group, and Gender</i>	189
Table 11.1	
<i>Students' Personal Views on the Benefits of Learning English as a Foreign Language</i>	199

List of Figures

<i>Figure 1.1</i>	The Location of Aceh Province within Indonesia	1
<i>Figure 1.2</i>	A Map of Aceh Province	2
<i>Figure 1.3</i>	An Ogive Curve	22
<i>Figure 2.1</i>	A Mythical View of the 'Ideal' or 'Full' bilingual	26
<i>Figure 2.2</i>	More realistic Conceptions of Bilingualism	26
<i>Figure 4.1</i>	Method Design Chart	88
<i>Figure 4.2</i>	Original Reading Comprehension Test Front Page	97
<i>Figure 4.3</i>	Revised Reading Comprehension Test Front Page	98
<i>Figure 5.1</i>	Item Category Curve for Item 1	115
<i>Figure 5.2</i>	Item Category Curve for Item 6	116
<i>Figure 5.3</i>	Item Characteristic Curve for Item 1	117
<i>Figure 5.4</i>	Item Characteristic Curve for Item 6	117
<i>Figure 5.5</i>	Person-Item Threshold Distribution for Reading Comprehension	118
<i>Figure 5.6</i>	Students' Reading Comprehension and Item Difficulties Calibrated on the Same Scale	119
<i>Figure 5.7</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 10	120
<i>Figure 5.8</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 12	120
<i>Figure 5.9</i>	Item Characteristic Curves by Gender for Reading Comprehension Item 1	121
<i>Figure 5.10</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 2	122
<i>Figure 5.11</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 8	122
<i>Figure 5.12</i>	Targeting of Reading Comprehension by Gender	123
<i>Figure 5.13</i>	Targeting of Reading Comprehension by Type of Language Instruction	124
<i>Figure 6.1</i>	Graph of Reading Comprehension by Group	131

<i>Figure 6.2</i>	Graph of Reading Comprehension by Gender	132
<i>Figure 6.3</i>	Graph of Reading Comprehension by Group (Bilingual/Monolingual) with Pretest as Covariate	134
<i>Figure 6.4</i>	Graph of Reading Comprehension by Gender with Pretest as Covariate	135
<i>Figure 7.1</i>	Curve for Item Category for Item 11	148
<i>Figure 7.2</i>	Item Characteristic Curve for Item 3	149
<i>Figure 7.3</i>	Item Characteristic Curve for Item 9 (Not Ideal but Satisfactory)	149
<i>Figure 7.4</i>	Targeting Graph: Student Measures in English Writing (upper-side) and Item Difficulties (lower-side) Calibrated on the Same Scale	150
<i>Figure 7.5</i>	Writing Map	151
<i>Figure 7.6</i>	Item Characteristic Curve by Gender for Writing Item 12	152
<i>Figure 7.7</i>	Item Characteristic Curve by Gender for Writing Item 3	152
<i>Figure 7.8</i>	Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 12	153
<i>Figure 7.9</i>	item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 3	154
<i>Figure 7.10</i>	Targeting of Writing by Gender	155
<i>Figure 7.11</i>	Targeting of Writing by Type of Language Instruction	155
<i>Figure 8.1</i>	Graph of Writing Test by Group	161
<i>Figure 8.2</i>	Graph of Writing Test by Gender	162
<i>Figure 8.3</i>	Graph of Writing Test by Group (Bilingual/monolingual) with Pretest as Covariate	164
<i>Figure 8.4</i>	Graph of Writing Test by Gender with Pretest as Covariate	165
<i>Figure 9.1</i>	Item Category Curve for Item 2	174
<i>Figure 9.2</i>	Item Characteristic Curve for Item 2	175
<i>Figure 9.3</i>	Person-Item Threshold Distribution for Behaviour	176
<i>Figure 9.4</i>	Behaviour Map	177
<i>Figure 9.5</i>	Item Characteristic Curves by Gender for Behaviour Item 26	178
<i>Figure 9.6</i>	Item Characteristic Curves by Gender for Behaviour Item 12	178
<i>Figure 9.7</i>	Item Characteristic Curves by Gender for Behaviour Item 38	179

<i>Figure 9.8</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 7.....	180
<i>Figure 9.9</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 2	180
<i>Figure 9.10</i>	Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 38	181
<i>Figure 9.11</i>	Targeting of Behaviour by Gender	182
<i>Figure 9.12</i>	Targeting of Behaviour by Type of Language Instruction	182
<i>Figure 10.1</i>	Graph of Behaviour by Group	187
<i>Figure 10.2</i>	Graph of Behaviour by Gender	188
<i>Figure 10.3</i>	Graph of Behaviour by Group (Bilingual/monolingual) with Pretest as Covariate	190
<i>Figure 10.4</i>	Graph of Behaviour by Gender with Pretest as Covariate	191
<i>Figure 11.1</i>	Students' Common Views on the Benefits of Learning English	201
<i>Figure 11.2</i>	Students' Confidence and Achievement.....	204
<i>Figure 11.3</i>	Learning English through the Media.....	205
<i>Figure 11.4</i>	Family Support in Learning English	207
<i>Figure 11.5</i>	Obstacles in Learning English	209
<i>Figure 11.6</i>	Other Views.....	210
<i>Figure 12.1</i>	Bilingualism as a Life Time Skill	226

List of Appendices

Appendix A	English Reading Comprehension Test	251
Appendix B	English Reading Comprehension's Ordered Scoring Scheme	255
Appendix C	English Writing Test	256
Appendix D	English Writing Scoring Rubric	258
Appendix E	Attitude/Behaviour Questionnaire	260
Appendix F	Original English Reading Comprehension Text	262
Appendix G	Original Attitude/Behaviour Test	264
Appendix H	Request Letter to Head of Provincial Education of Aceh	265
Appendix I	Consent Letter to Headmaster	266
Appendix J	Consent Letter to English Teacher	267
Appendix K	Consent Letter to Student (English Reading Comprehension)	268
Appendix L	Consent Letter to Student (English Writing)	269
Appendix M	Consent Letter to Student (Attitude/Behaviour Questionnaire)	270
Appendix N	Student's Sample Answer for Reading Comprehension	271
Appendix O	Student's Sample Answer for English Writing (Bilingual Student)	272
Appendix P	Student's Sample Answer for English Writing (Monolingual Student)	274
Appendix Q	Student's Sample Answer for Attitude/Behaviour Questionnaire (Bilingual Student)	276
Appendix R	Student's Sample Answer for Attitude/Behaviour Questionnaire (Monolingual Student)	277
Appendix S	Complete Students' Written Comments.....	278

CHAPTER ONE

INTRODUCTION

Background

This chapter introduces the reader to the Indonesian education system and to bilingual and monolingual education in Aceh province, Indonesia, and to the rationale and background behind the study, and thus to its aims.

Aceh Province

Aceh (pronounced ä'chā), is located on the northern tip of Sumatra island, bordered on the north by the Malacca Strait and on the south by the Indian Ocean. Its population reached 4,144,500 in 2002, with an area of 57,365.57 square kilometers (Kadin Aceh, nd). The provincial capital is Banda Aceh, which had a population before the tsunami of 264,618 and after the tsunami 203,553 (Nurdin, 2006). The first part of its name comes from the Persian *bandar* and means "port" or "haven" (Washington, 2010). Banda Aceh is known as 'Mecca's Verandah' (or 'doorway to Mecca'), because historically it has been a stopping place for Muslim pilgrims journeying by ship from the east to Mecca and as Islam entered the Indonesian Archipelago through Aceh. Figure 1.1 shows the location of Aceh Province within Indonesia, and Figure 1.2 is a map of Aceh Province.



Figure 1.1 *The Location of Aceh Province within Indonesia*



Figure 1.2 *The Map of Aceh Province*

Having been previously given five different names: *Aceh Lhee Sagoe*, *Seuramoe Makkah*, *Nangroe Aceh Darussalam*, *Banda(r) Aceh* and *Kutaraja*, Aceh province is rich in resources, mainly oil and natural gas. The most important agricultural products are rice, coconuts, rubber, tea and coffee. Due to its location, Aceh has been one of the first points of contact for foreigners in Sumatra. Indian, European and Arabic visitors have all had great influence through the spreading of religion and culture. Over different periods of time, its location by the Malacca strait has also led to intensive trade and high economic growth. Buddhism came to Aceh with Indian traders in the 7th century, while Islam was introduced by Arabic and Indian Muslims in the 13th century. A number of Muslim kingdoms and sultanates were subsequently established in the region. Aceh (formerly called Achin) reached the peak of its power in the early 17th century. The Dutch gained control of the coast in 1873 and engaged in a partly successful effort to subdue the interior until 1910 (Answer.Com, 2010). The Acehnese, like most Indonesians, are Muslim, but are generally more conservative. Aceh is today considered as the province in Indonesia where Islam has the strongest position.

Tsunami Aftermath

Until December 26, 2004, Banda Aceh was not frequently the subject of international discussion. On that day, a giant earthquake of 9.3 on the Richter scale occurred off the coast of Aceh province. The earthquake started a huge tsunami, sending massive tidal waves towards the coast, which destroyed a large number of coastal areas

in Aceh, Indonesia and numerous islands in the province of North Sumatra. The tsunami buried parts of the cities of Meulaboh and Banda Aceh (Community webs, 2010). It was the worst hit area out of all the locations hit: 220,000 people were estimated to have died in January 2005; and by 2009, 225,000 were estimated to have died (Education, 2010). According to the country's National Disaster Relief Coordination Agency, another 37,063 are still missing. In addition, the UN estimates that 655,000 people were homeless and sheltering in scattered refugee camps across the province (BBC, 2005). The highest tolls were among the women and children who were in the low lying coastal areas while their husbands were at sea fishing. Over 800 kilometres of coast were severely affected, often up to five kilometres inland. At least 654 villages were damaged or destroyed, more than 500,000 people lost their homes, and more than 150,000 children were left without schools (Education, 2010).

The Indonesian government, supported by massive financing provided by national and international sources, responded quickly, providing food, personnel and aid. In the first days, weeks and months following the disaster they were focused on providing emergency relief necessities such as food, clean water and shelter, and the resources needed to treat victims, stop the spread of disease, clear debris from roads and restore essential services. The swift and well-coordinated emergency effort managed to prevent a major outbreak of disease and hunger (ADB, 2009; Education, 2010).

Reconstruction and rehabilitation in Aceh were based on emergency relief demands and implemented with money, goods and services provided by the Indonesian government, the United Nations and extensive international support. This has resulted in major results and improvements. Total commitment, as of 31 December 2008, was \$6.8 billion and this has exceeded the minimum estimated to re-build to pre-tsunami levels (estimated at \$6.2 billion) (ADB, 2009). As of December 2008, a \$2.4 billion fund was used for housing and transport, and \$1.7 billion was used for health, education, and community-based development activities. Regarding the education sector, 395 schools covering all 23 districts in Aceh have received support for teaching and learning materials, books, computers, and repairs for libraries (ADB, 2009).

Education in Aceh Province

In reconstruction and rehabilitation in the educational sector, the Acehnese have decided to proceed with an Indonesian policy. As part of Indonesia, the Aceh government is entitled to run a national policy. The difference between the education systems in Aceh province from that of other provinces is due to the education privilege given to Aceh, in that Islamic wisdoms are added as local content for primary and secondary education.

The National Education System of Indonesia is generally aimed at elevating the intellectual life of the nation and developing the Indonesian people fully, that is, as people who are devoted to God, have good knowledge and high levels of skill, are in good physical and spiritual health, are independent and fair, and feel responsible for their countrymen and nation. The education system also strives to create patriotic spirit, strengthen love for the fatherland, enhance social solidarity and awareness of the nation's history regarding its heroes, and create a forward-looking attitude. The learning and teaching climate has to generate self-confidence and a learning culture among all layers of society that induces an attitude and behaviour of creativity, innovative thinking, and orientation toward the future. The education system has its roots in the Indonesian culture, based on *Pancasila* (the five official philosophical foundations of Indonesia) and the 1945 Constitution. Law No. 2/1989 concerns the National Education System, the system aims at generating abilities and increasing the standard of living and dignity of the Indonesian people in order to achieve the national development objectives (Menanet, n.d).

The entire national curriculum of Indonesia is designed around two basic parts, the Pancasila and the 1945 Constitution. Based on the Indonesian constitution, the goal of education is to prepare its citizens (and students) to develop creatively and emotionally and acquire the skills, knowledge, attitudes and spiritual values necessary for responsible, active, autonomous, productive, and democratic citizenship (Syahdan, 2009). The constitution also notes that education in Indonesia is divided into two major parts, formal and non-formal. A formal education is standard education obtained through an accredited source which is divided into three levels, primary, secondary and tertiary education (see Table 1.1). A non-formal education is usually run at pre-school

age and primary level, such as TPA abbreviation for Taman Pendidikan Al-Quran (Centre for Al-Quran Learning) which is usually held in mosques, Sunday schools which are usually held by churches, and music tuition, English tuition, archery classes, and so on that are held by various providers.

Primary education consists of early childhood and elementary school. From birth until the age of three, Indonesian children do not generally have access to formal education. From the age of three to four or five, they attend kindergarten (*Taman Kanak-kanak*). This education is not compulsory for Indonesian citizens, as the aim of kindergarten is to prepare them for primary school. The majority of kindergartens are private schools and 99.35% of the kindergartens (i.e. 49,000) in Indonesia are privately operated. The kindergarten years are usually divided into "Class A" and "Class B", students spending a year in each class. Children aged 6–12 attend *Sekolah Dasar* (SD) (literally Elementary School). This level of education is compulsory for all Indonesian citizens, based on the national constitution. In contrast to the majority of privately run kindergartens, most elementary schools are government-operated public schools, accounting for 93% of all elementary schools in Indonesia. Similar to education systems in the U.S.A and Australia, students must study for six years to complete this level. Some schools offer an accelerated learning program, where students who perform well can finish elementary school in five years (Exchange., n.d)

Secondary Education in Aceh consists of middle school and high school. Middle School, generally known by the abbreviation "SMP" (*Sekolah Menengah Pertama*) is part of primary education in Indonesia. Alternatively students can enroll in *Madrasah Tsanawiyah*, abbreviated to "MTs". SMP and MTs have a slightly different curriculum; with MTs has additional subjects on Islamic studies. After graduating from elementary school, students attend Middle School for three years from the age of 12-14. After three years of schooling and graduation, students may move onto High School, or cease formal education. There are around 22,000 schools in Indonesia, with a balanced ownership between the public and private sectors. In Indonesia, high school is generally known by the abbreviation 'SMA' (*Sekolah Menengah Atas*) and 'SMU' (*Sekolah Menengah Umum*)/'SMK' (*Sekolah Menengah Kejuruan*). SMA/SMU differs from SMK in its curriculum studies and student destinations. The students at SMU are prepared to advance to tertiary education while students of SMK are prepared to be

ready to join the workforce after finishing their school without going to university or college. Alternatively students can enroll in *Madrasah Aliyah*, abbreviated to “MA”. SMA/SMU and MA have a slightly different curriculum; with MA has additional subjects on Islamic studies. Based on the national constitution, Indonesian citizens do not have to attend high school, as the citizens are only required by law to attend school for nine years of education. This is reflected in the relatively low number of high schools in Indonesia, just slightly below 9,000 schools.

Tertiary education occurs after graduation from high school or college. Students may attend a university or higher education institution. Higher education institutions are categorized into two types, public and private, both of which are supervised by the Department of National Education. There are three types of higher education institution: Universities, Institutes and Academies or colleges (Mohandas, n.d).

Table 1.1
Formal Education of Indonesia

LEVEL OF EDUCATION	Year	Age
	Kindergarten	
	Playgroup	4
	Group A	5
	Group B	6
	Primary School (<i>Sekolah Dasar</i> or SD, <i>Madrasah Ibtidaiyah</i> or MI)	
PRIMARY EDUCATION	Year 1	7
	Year 2	8
	Year 3	9
	Year 4	10
	Year 5	11
	Year 6	12
	Middle School (<i>Sekolah Menengah Pertama</i> or SMP, <i>Madrasah Tsanawiyah</i> or MTs)	
SECONDARY EDUCATION	Year 7	13
	Year 8	14
	Year 9	15
	High School / Vocational School (<i>Sekolah Menengah Atas</i> or SMA and <i>Madrasah Aliyah</i> or MA / <i>Sekolah Kejuruan</i> or MK)	

SECONDARY EDUCATION	Year 10	16
	Year 11	17
	Year 12	18
Academy/Institute/Polytechnic/College/ University		
TERTIARY EDUCATION	Bachelor	Various ages (approx. 4 years)
	Master	Various ages (approx. 2 years)
	Doctorate	Various ages (approx. 4 years)

Indonesia has a national education system which controls the education system across the whole archipelago, including Aceh. This system was regulated under Law No.2/1989, and then changed to Law No. 20/2003 (Wijaya, 2008). When the education system was under Law No. 2/1989, the curriculum implemented was Curriculum 1994, which focus was content oriented. The grading system focused more on the cognitive side of student learning. The learners had to study in class for 40 hours every week. The syllabi used were under centralised control and all schools had to use the same syllabi. The focus of English language teaching in Curriculum 1994 was grammar and reading and the teachers were the source of information (Elvyanti, n.d).

After being used for ten years, curriculum 1994 was changed into Curriculum 2004, which was called the competence-based curriculum (*Kurikulum Berbasis Kompetensi* or KBK) (Elvyanti, n.d). Curriculum 2004 was under Law No. 20/2003. In Curriculum 2004, the focuses were product and competence. The concepts of learning were learning to know, learning to do, learning to live together and learning to be. The grading system combined the cognitive side, the affective side and the psychomotor side all together. The learning time was reduced from 40 hours to 32 hours every week, although the number of subjects taught was the same as Curriculum 1994. In the 2004 curriculum, the government gave freedom to teachers to create the syllabi which could fulfill the needs of the local area and the learners (Elvyanti, n.d), including those students in Aceh province.

However, Curriculum 2004 is now not used any longer. It has been changed into Curriculum 2006, but is still the curriculum under Law No. 20/2003, and is a further development of Curriculum 2004. In the implementation of this 2006 curriculum, schools are given the authority to design the syllabus, the learning hours, the academic

calendar and the standard competence of graduates (Elvyanti, n.d). Moreover, there are more parties – societies, parents and school committees – involved in the syllabus design.

Curriculum 2006 has been implemented by primary and secondary levels of education. The procedures of curriculum implementation are regulated under Law No. 20/2003. Meanwhile, the tertiary level of education is to implement Curriculum 2004. The curriculum implementation at the tertiary level is regulated under Law No. 20/2003 and the Decree of Minister of National Education No. 232/U/2000 (Nasional, 2001). This curriculum implementation means there are curriculum differences in the primary, secondary and tertiary levels of education which affect the teaching of English and the English learning process.

Just this year (2013) the new Curriculum 2013 has been implemented. The new feature of the curriculum is the banning of bilingual programs. This means reducing English teaching learning lessons.

In Aceh Province, as well as in other part of Indonesia, English has been taught as a foreign language since 1950 (Yuwono, 2005) and English has been a focus in the curriculum as a compulsory subject from the first year of Middle School (students 12/13 years old) up to the final year of High School (students 17/18 years old), over six years of secondary schools. However, some elementary schools (similar to Primary Schools in the Australian context, with students aged from 6-12), have put English into their school local content. Generally, English is not compulsory in elementary schools.

English is taught as a separate subject and it does not become the medium of instruction at any particular point in schooling. The majority of teachers are Acehnese-Indonesians with diplomas in teaching English as a foreign or second language, and the language is not taught by native speakers or expatriates. English is generally the same way in all schools regardless of the impact of the new curriculum, which now gives the right for schools to design their own content so that it best suits their school's needs.

The idea of bilingualism is not new to the majority of Indonesians. Most of the people in Indonesia are bilingual at an early age. They speak a local native language with their family whereas the official language is Bahasa Indonesia (also known as

Indonesian). As a language of instruction at school, Bahasa Indonesia is often the only language spoken. However, in remote areas throughout Indonesia, teachers need to alternate their teaching with the local native languages. The local native language is not dying out, even though it serves only as a spoken language.

Formal bilingual education is still seen as a ‘fantasy’ and a ‘luxury’ for the majority of Indonesian parents. Those parents who have good incomes mostly enrol their child or children into bilingual schools, which are usually international schools or private schools, while those parents with low income mostly enroll their child or children into state-run monolingual schools (Anecdotal evidence from experience, author Syahabuddin 2010).

Acehnese people share the same problem with that of the majority of Indonesian parents regarding bilingual education: that is, most parents want their children to be taught bilingually but there are insufficient bilingual schools and teachers. In the context of bilingual education in Aceh, numbers of private schools have been occasionally providing bilingual programs for some decades. It is important to bear in mind that the levels of bilingual programs offered vary from one private school to another. Although these schools do not claim schools with bilingual programs, one can argue that the programs they provide are similar to Developmental Bilingual Education. Just after the 2004 tsunami, when Aceh received support and help with education from international agencies, some international-funded bilingual schools suggested the idea of bilingual education for the re-building after the 2004 tsunami. Starting from that time people have become familiar with the terms ‘bilingualism’ and ‘bilingual education’ and with what it means to enrol their children into one of these schools. One famous and favourite bilingual school in Aceh is Fatih Bilingual School, which has operated since December 26, 2006 (Forum, 2006). Apart from Fatih Bilingual schools in Aceh, there are state schools that provided bilingual programs. Those schools are called International Standard Schools (also known as Sekolah Berstandar Internasional, or RBI) and Candidate to International Standard Schools (Rintisan Sekolah Berstandar Internasional or RSBI). Those schools returned to normal standard schools (with no bilingual programs) when bilingual education was banned in Indonesia in 2013.

In general, the context of bilingual education in Indonesia and Aceh is argued to be different from elsewhere. In Aceh, the aim for students learning English within a bilingual school is not to prepare them for a mainstream class (for example, an English mainstream); rather, the aim is to prepare them to be able to function in two languages, that is, Bahasa Indonesia and English. This function is said to be needed for their future academic studies and jobs. It is suggested that the context of bilingual programs in Aceh tends to be what it is called ‘Bilingual education in majority languages’ (Baker, 2011), where a strong language or a national language of the region is used alongside English as the medium of instruction.

During the last two years since the re-building after the 2004 tsunami, people in Aceh have noted that students graduating from bilingual programs have outperformed students graduating from monolingual state-programs. (This is anecdotal evidence and it is neither supported nor refuted by any reputable studies in Aceh). A question arises: In Aceh, are students learning in bilingual programmes performing and achieving better than students learning in state-run monolingual programmes? Based on this question, the present study will conduct research on bilingual and monolingual students’ achievement in English writing, English comprehension, and in attitude towards and behaviour surrounding learning English; and this study will also investigate teachers’ attitudes and behaviour regarding their bilingual and monolingual students’ ability in English.

Bilingual and Monolingual Education

This study is concerned with the teaching of English in both bilingual and monolingual schools and so the background focuses mainly on English instruction. The published literature, however, is not so finely focused and so the present review includes some more general research relating to bilingual and monolingual instruction.

Bilingual education, in broad sense, is termed as the use of two languages as a means of teaching and learning instruction (Anderson & Boyer, in Romaine, 1995) of selected subjects (Margana, 2009). More specifically, bilingual education is defined as education that aims to promote bilingual competence by using both languages as the media of instruction with respect to three features: linguistic goals, pedagogical

approaches, and levels of schooling, for significant portions of the academic curriculum (Genesee, 1987; Met, 1998).

In Aceh, as in other provinces in Indonesia, bilingual education aims to promote the maximum acquisition of English due to changes in the Indonesian educational context, which were the result of the increased role of English as an international language and the language of technology (Margana, 2009). In practice, most bilingual classroom settings use English as the medium of instruction, whereas Indonesian is used as the medium of instruction in all monolingual classroom settings. The monolingual classroom setting in Aceh, on the other hands, involves classrooms where the medium of instruction to explain English words, sentences, stories and context is Bahasa Indonesia (Classroom Observation, done by Khairiah Syahabuddin, 17 January 2011).

There are a number of reviews of the literature relating to bilingual education (see for example Genesee, 1987; Met, 1998; Rossell & Baker, 1996). Rossell and Baker (1996) stated that “of 300 program evaluations read, only 72 (25%) were methodologically acceptable” (p.7). They also found that “on standardized achievement tests, transitional bilingual education [i.e. a bilingual education type which teaches students in their native languages as a preparation for main stream English-only classrooms] is better than regular classroom instruction in only 22%” of the programs. So the question of which program is better, bilingual or monolingual, is open to question, as it is in Banda Aceh. Another major problem with all these studies is that the tests were done with scores based on True Score Theory. The items are not ordered by difficulty to form a scale and the scores based on them are non-linear, and often not unidimensional either (see Wright, 1999). Michell (1990, 1999), a measurement expert, claims that with True Score Theory, one cannot even claim that a student with a higher percentage score has more achievement than a student with a lower percentage score, by logic. In True Score Theory, a typical standard error of measurement (SE) for percentage scores with a split-half reliability of 0.75 and a standard deviation of 20 is $SE = 20 \sqrt{1-0.75} = 10$. That is, a student with a score of 80% does not necessarily have a better achievement than a student with a score of say 70%, within the error of measurement.

A problem with many of the published long term studies (such as over one year or more) is making a true causal inference when so many extraneous variables are not controlled between pretest and posttest. This means that the true cause of any alleged improved achievement is probably a combination of variables, such as differences in time-on-task, differences in English homework, differences in English after-school-hours television watched and English books read, not just a difference between school bilingual instruction or school monolingual instruction. This leads to the proposed study in Banda Aceh and what is needed to be done. The proposed study needs to use the best available measurement procedure (namely Rasch measurement), and the best available Rasch computer program (Rasch Undimensional Measurement Model, RUMM) (Andrich, et al., 2010), and it needs to control for any extraneous variables that may affect the student outcomes in writing, reading comprehension and attitude/behaviour between pretest and posttest.

Rationale

Aceh is unique as one of 33 provinces in Indonesia, because it has been given privileges which have not been given to any other provinces in Indonesia (Government, 2000). Aceh province has been given three unique privileges in education, religion, and culture. This means that the province has freedom to manage its governance in terms of education, religion and culture in order to meet its needs. In line with the privileges in education, Aceh has added Islamic wisdoms as local content under Indonesia's curriculum for primary and secondary education. For tertiary education, Aceh has allocated a great amount of funding for education, such as scholarship provisions for Acehese students to study abroad. In addition to that, at present, the Acehese government welcomes any discourse and thoughts to enrich Acehese education. This study will inform the Acehese government and Acehese people of any possible benefits of bilingual education in Aceh. Should the Aceh government agree on the idea of implementing a compulsory bilingual program into the education curriculum, it can pass on the information on bilingual education to state schools. It is not that the Indonesian government should base its policy on this one study but, if the present study shows a positive outcome, then further studies should be implemented to help develop education policy in Aceh.

While it is commonly believed by people in Banda Aceh that bilingual schools are better than monolingual schools, there is no clear evidence for this with Banda Aceh data. The students begin learning English in Year 7 and the common belief is that bilingually taught students are better at English writing, better at English reading comprehension, and have a better attitude and behaviour with regard to learning English than their counterparts in schools where teaching is monolingual. To answer their beliefs with evidence, this study compares bilinguals and monolinguals on Rasch-created linear measures of the three English outcomes over a two month period.

Statement of Purpose

The purpose of the present study is three-fold. The first is to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English text writing, English reading comprehension, and attitude and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual/monolingual). The second is to investigate the students' attitudes and behaviour in relation to the use of languages of instruction (combined English and Bahasa Indonesia vs. Bahasa Indonesia only) on their ability in English. The third is to investigate the students' perceptions about their learning English as a foreign language through their written comments.

Research Questions

1. Do first year middle state-school bilinguals in Banda Aceh achieve better in English Reading comprehension and English text writing than those of monolinguals?
2. Do first year middle state school bilingually-students at bilingual schools in Banda Aceh have better attitudes and behaviours with regard to learning English than those at monolingual schools?
3. What are the attitudes and behaviours of first-year middle state school bilinguals with regard to bilingual and monolingual education in Banda Aceh in terms of learning English?
4. What are first-year middle state school students' perceptions with regard to their learning of English as a foreign language?

Significance

The study is important for the following four reasons. One, there have been no previous studies comparing bilingual and monolingual educational outcomes with Aceh data or with any data in a Muslim province of Indonesia. Since the re-building of the education system in Aceh after the 2004 tsunami, the people of Aceh are asking the question about which type of instruction is better and the present study will thus provide data in relation to answering this question.

Two, the study has sufficient sample size and sufficient power in both the control group (N=386) and the experimental group (N=394) to give a definitive answer to the research questions. There are adequate controls and monitoring for extraneous variables such as time-on-task, content taught, English homework given, watching English television after school hours, reading English books after school hours, and after-school English learning. Consequently, the present study has the capacity to provide a comparison result that is credible.

Three, there have been no previous similar studies done with linear, unidimensional Rasch measures comparing bilingual and monolingual educational outcomes and all known previous studies in other countries comparing bilingual and monolingual educational outcomes have used non-linear scores based on True Score Theory measurement. The present study thus does not only provide good quality data to answer the research questions but it also uses current world's best measurement practice to produce linear, unidimensional measures of three educational outcomes (English writing, English reading comprehension, and attitudes and behaviour with regard to learning English).

Four, the present study should give direction for some future research into bilingual and monolingual education that should lead to good policy development in Aceh. The present study has clear educational policy implications. If the present study shows that there is a clear advantage in educational outcomes for bilingual or monolingual education, then this will 'drive' further research and provide important information for education decision-makers in Aceh. The present study can be regarded as an important future foundation for the educational system in Aceh on how to lead English teaching and learning in Aceh, especially in bilingual education.

Limitations

There are four main limitations to the present study. One, strictly, the data and the results only apply to Aceh province and not to any other Muslim provinces in Indonesia. Even though Aceh province has some similarities to other Muslim provinces in Indonesia, there are some substantial differences in other provinces, such as being more secular; these differences may make differences in educational outcomes possible. On the other hand, it might be assumed that the results applicable to Aceh province should be 'more-or-less' applicable to other Indonesian provinces, until proven otherwise.

Two, particularly, the results of the present study are not applicable to other Muslim countries because Aceh province has a unique cultural and trading background, and because of the redevelopment of the educational system following the 2004 tsunami. Other Muslim countries, especially in the middle-east, have long histories of Muslim economic development, education and culture that are different from Aceh and their educational systems are different. The differences might make for some differences in the way that bilingual and monolingual education occurs and hence for differences in educational outcomes.

Three, exclusively, the data for the present study only apply to Year 7 of middle school and not to any other years. The results will be generalizable to other Year 7 students in Aceh, Indonesia, but not to other schooling years in Aceh, Indonesia, because the tests were conducted for the Year 7 students only, not to other schooling years' students. It might, however, be expected that there will be little differences across other years. If the results are true for Year 7, then they are likely to be true for Years 8-12, but only research can tell.

The reason of choosing the Year 7 (semester 2) students for the study as the respondents was because they were beginners in learning English, and therefore, were assumed to share a similar starting point in learning English, and any experience gained during the study could then be associated with the medium of instruction. Under Curriculum 2006, English is taught as part of the curriculum in Year 7 semester 1. The present study investigated students' English ability and behaviours with regards to language of instruction. Year 7 (semester 2) students were considered suitable for the

study because they were beginners. They just had learned English in that schooling year (Year 7) and just had been learning English for 5 months prior to the study. They shared similar abilities in English and were considered to have a similar novice command in English. Within the two months of participating in the study, the two groups of students (Bilingual and Monolingual) were more likely to experience in-class English learning differently, due to the view that the language of instruction was different. Students learning English in a bilingual program were taught using combined English and Bahasa Indonesian. Students learning English in a Monolingual program were taught with the majority of Bahasa Indonesia and small amount of English. If other year students (for example, Year 8 students) were taken as the respondents of the study, they were not beginners anymore in learning English and the language of instruction might not be the only difference between them in regards to their English ability and behaviours because they would have had differing relative amounts of English knowledge and ability, and perhaps very different English learning behaviours. The two month difference between pretest and posttest may not have sufficient effect on them.

The present study used pretests and posttests separated by a two-month period and, while this should be enough time for differences in educational outcomes to occur, it may be that time differences over longer periods such as one or two years would produce larger differences. A problem for larger time-span research is that add-on variables such as after-school differences in time-on-task and after-school differences in exposure to English writing, speaking, reading, and watching television and movies may be just that or they may be due to the bilingual program. It is difficult to administer experiments to determine ‘what causes what’ here.

Four, solely, the data only apply to two English skills, namely, English Writing and English Reading Comprehension and not to the other two English skills, namely, English listening and English speaking. Further research is needed with a similar format and linear measures to determine the result for English listening and speaking.

Definition of Terms

The following definitions explain the meaning of the terms as used in the context of the present study.

Bilingualism

The term 'bilingualism' is reserved to describe two languages used by individual students, regardless of students' fluency in each of the languages.

Bilingual Education

Bilingual education is defined as education that aims to promote bilingual competence by using both languages as media of instruction with respect to three features: linguistic goals, pedagogical approaches, and levels of schooling, for significant portions of the academic curriculum. More specifically, for the purpose of the present study, it is defined as the teaching of English by using a combined English and Bahasa Indonesia with equal percentage of time.

Bilingual Classroom Setting

A bilingual classroom setting is an English language classroom in which English is used as the medium of instruction, whereas Indonesian is used as the medium of instruction in all other classes.

Monolingual Education

A monolingual classroom setting is an English language classroom in which Bahasa Indonesia is used as the medium of instruction to explain English words, sentences, stories and context. More specifically, for the purpose of the present study, it is defined as the teaching of English by using a majority of Bahasa Indonesia and a small amount of English by time.

English Reading Comprehension

English Reading Comprehension is the ability of second language readers to obtain meaning from texts by actively using both lower and higher skills to decode the smaller elements and construct the meaning. By using their schemata (previous or background knowledge), they are able to understand the main idea, sequence the order, and obtain detailed information. The English reading comprehension in this study represents student English achievement and it is based on the Bloom Taxonomy.

English Reading Comprehension Test

The English Reading Comprehension Test used in the present study consists of 12 multiple choice items and three written items to be answered by the students after reading a given piece of English text (see the full test in Chapter Three). A special scoring rubric in which scoring is ordered by item difficulty for the multiple choice items and ordered by quality for the written answers was designed for the test (see the scoring rubric in Chapter Three). This scoring is consistent with Rasch measurement principles and is used to create a linear, unidimensional measure for the variable English Reading Comprehension.

English Text Writing

English Text Writing is the ability of second language writers to produce written texts by actively using writing strategies, techniques and skills, which have been acquired during their English language learning instruction, whether it be in bilingual or monolingual classes. The English text writing in this study represents student English achievement and it is based on Bloom's Taxonomy (Bloom et al., 1956).

English Text Writing Test

The English Paragraph Writing Test consists of two compulsory topics on which the students are asked to write several paragraphs in English (see the full test in Chapter Three). A special scoring rubric in which scoring on three aspects of writing is ordered by quality was designed for this test (see Chapter Three). This scoring is in line with

Rasch measurement principles and is used to create a linear unidimensional measure of English Writing Quality.

Attitude and Behaviour with regard to Learning English

The Attitude and Behaviour with regard to Learning English Questionnaire (see Appendix E) contains 18 items: three on tasks for listening, three on tasks for speaking, three on tasks for reading, three on tasks for writing, three on tasks for student/student relationships, and three on tasks for student/teacher relationships. Each item was answered from the two perspectives of attitude and behaviour, that is ‘ideally, this is what I think should happen’ (attitude and easier); and ‘this is what actually happened’ (behaviour and harder). The full questionnaire is given in Chapter Three. Response category ‘most or all of the time’ was scored 3, response category ‘some of the time’ was scored 2, and response category ‘never or rarely’ was scored 1. This scoring is ordered in line with Rasch measurement principles and was used to create a linear unidimensional measure.

Middle School

Middle school in Aceh is the level after secondary school. Middle school students are 12-15 years old. The duration of middle school is three years, and English classes are begun in the first semester of the first year of this level.

True Score Theory Measurement

True Score Theory is a way of measuring variables in the social sciences (and education) which claims that the observed total score obtained by a person on a set of test or questionnaire items is made up of a ‘true score’ and a random error score. The scale created by True Score Theory does not contain equal units of measures and is therefore non-linear. That is, the difference between, for example, 50% and 60% does not represent the same amount of variable difference as between 70% and 80%. True Score Theory scores are commonly considered to have at least six problems: (1) non-linearity; (2) multi-dimensional with ‘noise’; (3) item difficulties not ordered; (4) person ‘measures’ and item difficulties not ordered on the same scale; (5) the ‘measures’ are test (item content) dependent; and (6) the ‘measures’ from different tests, even on the

same topic, cannot be validly added or linked onto a single scale (see Michel, 1990, 1999; Smith, 1996; Waugh & Chapman, 2005) .

Rasch Measurement

In Rasch measurement, items are ordered from easy to hard on a continuum and their difficulties are calculated on a linear scale (a log odds scale). The person measures are calculated on the same linear scale. An important point to understand is that when the data fit a Rasch measurement model, the differences between the person measures and the item difficulties can be calibrated together in such a way that they are freed from the distributional properties of the incidental parameter, because of the mathematics involved in the measurement model. This means that ‘scale-free’ measures and ‘sample-free’ item difficulties can be estimated with the creation of a mathematically objective linear scale with standard units. The standard units are called logits (the log odds of successfully answering the items) (This has been taken from Waugh, 2003, 2005, 2010a; Waugh, 2010b).

Unidimensionality of Variables

In the present study, unidimensional measures are created for English Writing, English reading Comprehension, and Attitude and Behaviour about Learning English. These measures involve a variety of aspects including low order thinking (such as knowing facts and basic comprehension), higher order thinking (such as analysis, synthesis and evaluation), low and high order attitudes, and physical dexterity and, in this sense, cannot be unidimensional. With Rasch measurement, unidimensional means that a single parameter for each person (person measure) can be created as applying to all of the scale items, that a single parameter can be created for each item (item difficulty) applying to all the persons measured on the same scale, and that these parameters can be applied to accurately predict each person’s response to each item.

Person Separation Index

Person Separation Index is an index ranging from 0 to 1 that shows the proportion of observed variance considered to be true. A high value of the index indicates that measures of the respondent’s ability or preference are sufficiently well

separated along the scale in relation to the errors of measurement. It is “structured as the ratio of estimated observed variance among persons, using estimates of their locations (measures) and the standard errors of these locations (measures)” (Andrich & van-Schoubroeck, 1989, p. 483) . The Person Separation Index is interpreted like Cronbach’s alpha which measures the internal reliability of non-linear scales (Cronbach, 1951).

Item Thresholds

Item thresholds show the location on a continuum whereby it is likely a person will obtain a particular score. More specifically, thresholds are points between adjacent response categories where the odds of answering in either category are 1:1. With three response categories there are two thresholds and with four response categories there are three thresholds. Thresholds should be ordered in line with the ordering of the response categories showing that the responses are answered consistently and logically (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Standardised Residual

Residuals are differences between the expected response according to the Rasch measurement model and the actual response. The standardized residual is the residual divided by its standard deviation. When there are many standardized residuals, then the mean should be close to zero and the standard deviation should be close to one, when the data fit the Rasch measurement model (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Response Category Curves

Response Category Curves show whether items have been answered logically and consistently. The actual curve that is produced shows the relationship between the probabilities of answering each category in relation to the specific measure. For example, the ideal curve for an item with three response categories shows that when the measure is low, the probability is high that the participant response is low (category one). As the measure increases, the probability of answering category one decreases and the probability of answering category two increases. As the measure increases further

still, the probability of answering category two decreases and the probability of answering category three increases (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

Differential Item Functioning

Differential Item Functioning refers to items that give different success rates for two or more groups, at the same ability level (Holland & Wainer, 1993). Masters (1988a) states that item bias occurs if an item's estimated difficulty is significantly greater when calibrated on one sub-group than when calibrated on the other, resulting in the item being considered 'biased' with respect to those two sub-groups. In other words, test bias can occur when the test requires different information or knowledge than that being tested, causing test scores to be less valid for a particular group of test-takers (see also Penfield & Lam, 2000).

Item Characteristics Curves

Item Characteristic Curves show how well the items differentiate between persons with differing measures. An ogive curve (see Figure 1.1 for an ogive curve) is produced for each item showing the relationship between the expected response score and the particular measure (Andrich, et al., 2010; *RUMM 2030 Manual*, 2009).

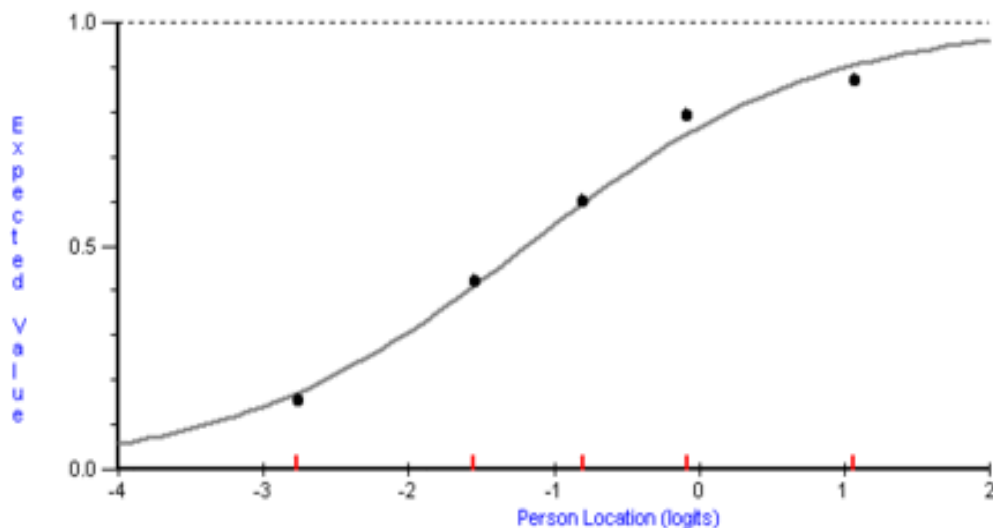


Figure 1.3 An Ogive Curve

Structure of the Thesis

This thesis is presented in twelve chapters. Brief summaries of the contents of Chapters Two to Twelve are given below.

Chapter Two is the literature review. The chapter provides a discussion on bilingualism and bilingual education in the world. It includes reviews of research results relating to bilingual and monolingual education and also the results of individual studies in western and Muslim countries. The research literature is critically analysed to show the gaps in the research and its deficiencies in implementation of research projects.

Chapter Three presents the conceptual framework of bilingual and monolingual teaching and it explains the expected outcomes according to modern theory. A discussion on the deficiencies of measurement using True Score Theory is provided and Rasch measurement is explained as a strong improvement in line with world's current best practice. The structure of the measurement instruments for the three variables to be used in the present study, that is, the English Writing Test, the English Reading Comprehension Test, and the Attitude and Behaviour Questionnaire about Learning English, are described, along with the scoring rubrics for each variable.

Chapter Four presents the mixed-method design. This involves the quasi-experimental approach of intact control and experimental groups with pretest and posttest measures, and their associated samples and piloting of questionnaires. The design involves Focus Groups and video recording of several bilingual and monolingual classrooms, with their associated samples and piloting. The discussion expands on the control of extraneous variables in the quasi-experiment, test data collection, data entry and data analysis.

Chapter Five (Data Analysis, Part 1) presents the Rasch data analysis results for English Reading Comprehension. The results of the RUMM2030 computer program output involving tabular data such as overall item and person fit, individual item and person fit, item-trait interaction (dimensionality), thresholds, and reliability (Person Separation Index) are presented and explained. Graphical output from the RUMM2030 program involving Response Category Curves, Item Characteristic Curves (ogives) and targeting graphs are also presented and explained. Similar RUMM2030 output results

for English Writing are presented in Chapter Seven (Data Analysis, Part 3) and for Attitude and Behaviour to Learning English in Chapter Nine (Data Analysis, Part 5).

Chapter Six (Data Analysis, Part 2) presents the ANCOVA and ANOVA results for the quasi-experiment interaction and main effects for English Reading Comprehension. Similar output results for English Writing are given in Chapter Eight (Data Analysis, Part 4), and output results for the Attitude/Behaviour Questionnaire are presented in Chapter Ten (Data Analysis, Part 6).

Chapter Eleven (Data Analysis, Part 7) presents the data analysis of the students' written comments.

Chapter Twelve presents the discussion and implications of the results of this study in the light of current literature findings. The chapter concludes with implications for bilingual education in Aceh province for teachers, students, policy makers and future research.

The next chapter reviews literature on bilingualism and bilingual education.

CHAPTER TWO

BILINGUALISM and BILINGUAL EDUCATION

There is a great deal of literature covering many years past on bilingualism and bilingual education in the relevant journals, books and reports. Clearly, bilingualism and bilingual education has been, and continues to be, an important and controversial topic in many countries across the world. There are too many studies and reports to summarise all of them for the present chapter of this study. This chapter briefly revisits the nature of bilingualism, the nature of bilingual education, and the nature of bilingual education in Indonesia and summarises some of the more relevant and recent research (2003-2013) on these topics from countries across the world. In this chapter, there is a selection and summary of research commentary on the nature of bilingualism, the benefits of bilingualism, the disadvantages of bilingualism, the nature of bilingual education, models of bilingual education, bilingual education in Indonesia and the recent 2013 judicial banning of bilingual teaching in Indonesia.

The Nature of Bilingualism

What does it mean to be a bilingual?

To define the term bilingual is not easy. It has been defined as one's knowledge and one's use of two languages (Butler, 2013). The term might mean that one's ability in the two languages is balanced. However, this is rarely the case, due to the fact that the level of one's bilingualism differs from one bilingual person to another. In regards to the four language skills (listening, speaking, reading, and writing), for example, a bilingual person might be strong at the two productive skills (speaking and writing), but might be weak at the other two receptive skills (listening and reading) (Baker, 2011), or *vice versa*. So, a question arises, "How bilingual is bilingual?" In line with this, Shih (2012) believed that 'a balanced bilingual' is a myth and referred to it in the form of circles by analogy (see Figure 2.1).

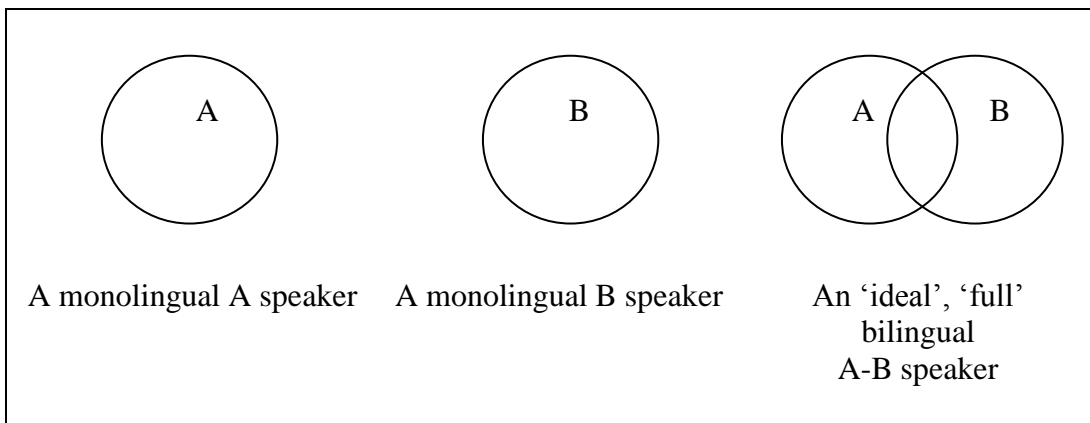


Figure 2.1 *A Mythical View of the 'Ideal' or 'Full' Bilingual (Source: Shih, 2012, p.4)*

The size of the circles in Figure 2.1 indicates, by analogy, that the level of proficiency in each language is the same for a bilingual speaker and for his/her bilingual ability. In reality, one's bilingual ability is rarely 'balanced'. Although a balanced bilingualism is an ideal concept, only a few bilingual speakers can attain it (Baker, 2011; Wei, 2000). More often, one's bilingual ability varies from another bilingual and so the circle analogy should be more like those depicted in Figure 2.2.

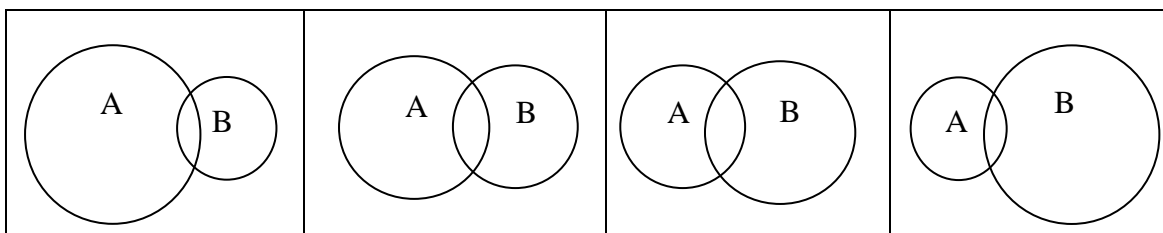


Figure 2.2 *More Realistic Conceptions of Bilingualism (Source: Shih, 2012 p. 5)*

Grech and McLeod (2012) have summarised the breadth of definitions of bilingualism into three categories:

1. Bilingual exposure from birth;
2. Using more than one language in day-to-day functioning; and
3. A continuum of use and proficiency in more than one language.

Due to the difficulty of representing a simple definition of bilingualism, Baker (2006, pp. 3-4) outlined eight dimensions of bilingualism. These dimensions are summarised and explained here due to their contribution towards the question posed earlier, 'How bilingual is bilingual'?

Eight Dimensions of Bilingualism summarised from Baker (2006, pp. 3-4):

1. Ability

Ability here refers to whether bilinguals use both languages actively or passively. Some have the ability to speak and write actively in both languages. This ability is called productive competence. Others, on the other hand, have the ability to listen and read passively in both languages. This ability is called receptive ability. Some bilinguals might only receive understanding from hearing and reading but they are not ready to produce the form of utterances (speaking ability) and written discourses (writing ability) for the two languages. Valdes et al. (2003), quoted in Baker (2006), affirm that ability is on a continuum. Bilinguals vary their ability with dominance of one language over the other, and vary in their development of one language over the other.

2. Use

The places and settings where the languages are obtained vary. Some bilinguals might acquire an additional language at home, school, on the street, and/or phone, Facebook, or Twitter. The languages are often used for different reasons or purposes.

3. Balance of two languages

Most often the use of the two languages is imbalanced, and one language is dominant over another.

4. Age

Bilinguals may acquire languages at different ages. Some acquire the two languages from birth. This is called infant or simultaneous bilingualism. Some acquire one language after the age of three. This is called sequential or consecutive bilingualism.

5. Development

The development of the languages varies. When a language is fully acquired, while the other has just started, it is called incipient. When the second language is

developing, the bilinguals are called ascendant. When one of the languages is decreasing, the bilinguals are called recessive bilinguals. When a language starts to stop, the bilinguals are called language attrition bilinguals.

6. Culture

If the person learning a language is also learning the culture of the language, the process is called acculturation. Bicultural competence tends to relate the knowledge of the two cultures of the two languages as well as the feelings, attitudes, and behaviours of the bilinguals.

7. Context

The places where the bilinguals live vary. Some bilinguals live in a bilingual/multilingual place where two or more languages are used daily (known as endogenous context). Some live in places where there are no second language communities (known as exogenous contexts). Some bilinguals have contact with second language communities through vacations, phone, email, Facebook, Twitter, television, films, books and so on (known as network contexts). Some contexts may be subtractive, where the home language is replaced by the community language due to the politics of the majority language users. Other contexts are additive, where a person adds a new language without any cost occurring; this is the case for some elite or prestigious bilinguals.

8. Characteristics

Elective bilingualism is a characteristic of people who prefer to learn a second language in a classroom. They add a second language without losing their first language. Circumstantial bilinguals learn another language to function effectively due to their life demands. For this characteristic, there is a concern at losing their first language, – a subtractive context. These two characteristics (elective and subtractive) are important, because they lead to practice, to status, to politics, and to power issues among bilinguals.

In line with the eight dimensions, Butler (2013) provided a concise table which is called ‘Classical Typology of Bilingualism’, and it is set out below (see Table 2.1).

Table 2.1
Typology of Bilingualism Classical

<i>Point of Focus (Dimension)</i>	<i>Typology</i>	<i>Definitions</i>	<i>Characteristics of Second Language Acquisition</i>	<i>Related issues and educational implications</i>	<i>Additional complications in multilingual contexts</i>
Relationship between proficiencies in two languages	Balanced; Dominant (Peal & Lambert, 1962)	Achieving equal level of proficiency in L2 and L1 (balanced); L2 proficiency varies but not the same as L1 (dominant)	Functional differences; related to age factor	Conceptualizing and assessing language proficiency; Cummins’ threshold hypothesis and interdependent hypothesis; Semilingualism	Greater complexity in conceptualizing and measuring multilingual competences
Functional ability	Receptive; Productive	Understand but not produce L2 either in oral and/or written domains (receptive); understand and produce L2 (productive)	Functional and motivational differences	Language use irrespective of proficiency levels and identity	Greater diversity in functional differences across domains and across languages
Age of acquisition	Early; Simultaneous; Sequential; Late (Genesee et al. 1978)	Exposed to two languages from birth (simultaneous); Exposed to L2 after L1 has some foundation (sequential); became bilinguals during adulthood (late)	Maturational difference; schooling differences	Neurolinguistic differences; critical period hypothesis	Greater diversity in the acquisition order, can have multiple L1s and/or L2s
Organisation of linguistic codes and meaning unit(s)	Compound; Coordinate; Subordinate (Weinreich, 1953)	Two sets of linguistic codes stored in one meaning unit (compound); stored separately (coordinate);	Functional differences; differences in form-meaning mapping	Difficulties with operationalising distinctions and testing differences	Greater complexity and diversity in multilingual memory organization according to typological differences

		L2 is assessed through L1 (subordinate)			among languages and proficiency levels
Language status and learning environment; literacy support of L1	Elite; Folk (Fishman, 1977); Circumstantial; Elective (Valdes and Figueroa, 1994)	No or little additive value of L1 as a language minority status (folk / circumstantial); additive value of L2 (elite/elective)	Differences in language status and value of bilingualism	Support for literacy in L1 and L2 literacy development	Greater diversity in social values attached to multiple languages
Effects of L2 learning on the retention of L1	Additive; Subtractive (Lambert, 1974)	L2 as enrichment without loss of L1 (additive); L1 is replaced by L2 (subtractive)	L2 as enrichment with or without loss of L1; status of a language in a given context	Social value of L1 greatly influences the retention of L1; support for literacy in L1 and L2 literacy development	Greater complexity of learning an additional language from previously acquired languages; greater diversity of status across languages
Cultural identity	Bicultural; L1 monocultural; L2 accultural; Deculturated (Hamers and Blanc, 2000)	Cultural identity shaped by two cultures (bicultural); identity in one culture (L1 monocultural); loss of L1 culture (L2 accultural); identity in neither cultures (deculturated)	Differences in acculturation process	High bilingual competence does not necessary coincide with dual identity	Multiple cultural identities coexist irrespective of competences

Source: (Shih, 2012, pp. 113-114)

This table gives us some more comprehensive ideas on what it means to be a bilingual, as there are numbers to consider, such as, point of focus (dimension), typology, definitions, characteristics of second language acquisition, related issues and educational implications, and additional complications in multilingual contexts.

To sum up on the terminology of bilingualism, the literature suggests that perhaps a better way to label a bilingual is to consider the dimensions: ability, use, balance of the two languages, age, development, culture, context, and characteristics.

What are the Benefits of Being a Bilingual?

It has been documented through a great number of research studies that people who are bilingual and/or have received bilingual education outperform people who are monolingual and/or have received monolingual education on cognitive matters.

It has been documented through a great number of research studies that bilingualism and bilingual education outperforms monolingualism and monolingual education on cognitive matters. Among them is executive control, metalinguistic awareness, working memory, metacognitive awareness, and divergent thinking. Cognition has been defined as the ability to act in line with goals on learning and the ability to resist inference Morton and Harper (2007). An increasing body of studies finds many positive effects of bilingualism on cognitive advantage and evidence has been found that bilingualism enhances the cognitive development of children in a variety of circumstances and cultures (see Bialystok, Craik, & Freedman, 2007); Diamond, 2002; Kave, Eyal, Shorek, & Cohen-Manfield, 2008).

Executive control

Particular evidence regarding cognitive development is on executive control functions. Executive control, which is believed to be a most important aspect of cognitive development (see Diamond, 2002), refers to a set of interrelated processes in the frontal lobes (Bialystok & Viswanathan, 2009). It consists of three components, namely, (1) shifting between tasks on mental sets; (2) updating and monitoring memory presentations; and (3) inhibition of dominant or pre-potent responses (Miyake, Friedman, Emerson, & Witzki, 2000). Cognitive development is needed so that the two languages [became] integrated with the linguistic circuits used for language processing, creating a more diffuse, more bilateral, and more efficient network that supports high levels of performance” (Bialystok, 2011, p. 233).

The effects of enhanced executive control develop earlier in bilingual children than in comparable monolinguals (Adi-Japha, Berberich-Artzi, & Lidnawi, 2010; Ellen

Bialystok, 2010; Carlson & Meltzoff, 2008) and have been found at all stages of the life span, beginning from infancy (see Bialystok, Craik, Klein, & Viswanahan, 2004; Carlson & Meltzoff, 2008; Costa, Hernandez, & Sebastian-Galles, 2008; Kovacs & Mehler, 2009; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011). Evidence for executive control comes from behavioural and imaging studies. Results of the behavioural studies indicate that interference occurs to participants' other language (Francis, 1999; Kroll & deGroot, 1997). Results of the imaging studies indicate that interference from the non-target language is important (Colome, 2001; Marian, Spivey, & Hirsh, 2003; Rodriguez-Fornells, Rotte, Heinze, Nosselt, & Munte, 2002; Wu & Thierry, 2010). These behavioural and imaging studies reveal that both languages of bilinguals are constantly active to some degree (Bialystok, 2011), and available when either of them is in use (Kroll, Bobb, & Wodniecka, 2006; Marian, et al., 2003; Rodriguez-Fornells, Balaguer, & Munte, 2006; Sumiya & Healy, 2004), even in strongly monolingual contexts. The brain controls any inhibiting attention and hinders intervention from one language to another in joint activation for two languages (Greenberg, Bellana, & Bialystok, 2013).

Metalinguistic awareness

Metalinguistic awareness has been defined as the ability to think about language in terms of its form and structure and the relation between the two to produce comprehensible meanings (Adesope, Lavin, Thompson, & Ungerleider, 2010). It can also be recognised as the ability to direct attention to the systematic characteristics of language, and an ability to mirror them (Lazaruk, 2007).

Research on metalinguistic awareness has been conducted by numbers of reseachers (Baker, 2006; Bialystok & Barac, 2012; Chow, McBride, Cheung, & Chow, 2008; Kieffer & Lesaux, 2012; Kirby, Desrochers, Roth, & Sandy, 2008; Laurent & Martinot, 2009; Lazaruk, 2007; Rauch, Naumann, & Jude, 2011; Rispen, McBride-Chang, & Reitsma, 2008). Studies on metalinguistic awareness have concentrated on three awarenesses: words awareness: phonological awareness (also known as sound system awareness) and syntactic awareness (also known as word-order awareness) (Chow, et al., 2008).

Research on word awareness has consistently shown that bilingual children retain an astonishingly good grasp of relations between words and languages and the meanings of the words in the two languages. An example is the study by Bialystok and Barac (2012). The aim of their study was to ascertain the nature of the experiences of bilingual learners that make a contribution to performance on metalinguistic awareness. Their results uniformly demonstrated that bilinguals' metalinguistic performance improved as the knowledge of both languages increased. This is in line with another study conducted by Laurent and Martinot (2009) on phonological awareness in children exposed to the early learning of a second language. It aimed to investigate whether this exposure to the second language highlights the learners' development of phonological awareness. They showed that bilingual learners show more significant positive results in their phonological awareness than do their monolingual counterparts (see also Laurent & Martinot, 2009). Regarding this ability, the study by Baker (2006) on metalinguistic awareness among bilingual learners supported the view that bilingual learners 'gratify' their heightened metalinguistic abilities, instead of only possessing a universal metalinguistic dominance over monolingual learners.

Working memory

Working memory is defined as a temporary storage of information to be used for performance of various types of tasks (Baddeley, 2003). It is also seen as memory at work, not just memory (Conway, Jarrold, Kane, Miyake, & Towse, 2007). A reasonable amount of research has been conducted on working memory (see Abu Rabia & Siegel, 2011; Biedron & Szczepaniak, 2012; Morales, Calvo, & Bialystok, 2013; Ransdell, Barbier, & Niit, 2006; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011). A study done by Biedron and Szczepaniak (2012) investigated the role of short-term memory and working memory in accomplished multilinguals. The analysis revealed that short-term memory and working memory abilities in the multilinguals were higher than in mainstream students. Just recently, Morales, et al. (2013) explored two studies comparing the performance of bilingual and monolingual learners on tasks requiring different levels of working memory. Both studies show that bilingual children received advantages over their counterparts regarding working memory.

Metacognitive awareness

Metacognitive awareness refers to the awareness a person has of their own knowledge with regard to their mental activities and learning strategies (Adesope, et al., 2010). The process of learning two languages, which includes two systems of vocabulary, syntax, phonology, and morphology, and the process of using them in meaningful and appropriate contexts, may tell the bilingual learners how they manage to use their two cognitive processes and learning strategies (Kemp, 2007). Over the past decade, research has been conducted on the effects of bilingualism on learners' metacognitive awareness. The majority has found that bilingual learners show greater cognitive awareness than the monolinguals do (see Judge, 2011; Ransdell, et al., 2006; Vorstman, Swart, Ceginskas, & Bergh, 2009).

Divergent thinking

Divergent thinking can be understood as one's mental process and methods which are used to trigger creativity with the help of various alternative results (Kharkhurin, 2008). Research that investigated the outcomes of bilingualism on a learner's divergent thinking shows that bilingual learners develop greater ability to trigger and process numerous unconnected notions (Ghonsooly & Shawqi, 2012; Kharkhurin, 2008, 2009).

Is there any Disadvantage to being Bilingual?

There have been concerns over possible disadvantages of bilingualism, especially in the early years of the practice of bilingualism in the USA. Early bilingualism was seen as dangerous, leading to language disorder and language delay. These concerns have been associated with a deficit viewpoint. This deficit viewpoint was based on the idea that a child's mind could only hold one language at the same time, that is, more than one language was too much to bear. That concern might have taken root from a quotation from a professor that is now seen as an historical and hysterical comment (Baker, 2011): "If it were possible for a child to live in two languages; at once equally well, so much the worse. His intellectual and spiritual growth would not thereby be doubled, but halved. Unity of mind and character would have great difficulty in asserting itself in such circumstances" (in Baker, 2011, p. 139). That

fear was believed by certain people who associated it with deficiencies in the size of vocabulary gained in the two languages; competence and performance of the two languages and language creations, meaning and imagery (Wei, 2010). However, this fear has proved to be incorrect (Baker, 2011). It is widely believed at the present that having more than one language does not bring any harm to a child's mental ability. In fact, much research has shown that bilingual children excel over monolingual children in several aspects of cognition, as already stated.

Added to the deficit viewpoint are another four fears that have been outlined by Baetens-Beardsmore (2003): parental fear, cultural fear, educational fear, and politico-ideological fear. Parental fear has been associated with the unreadiness of parents to allow their children to encounter bilingualism. This unreadiness could be due to their own uncomfortable experience of being an adult trying to learn a second or foreign language. This fear has also been associated with misguided information received by parents from educational 'experts' who seem to believe in the deficit viewpoint. This parental fear could hinder their children's encounter with bilingualism.

Cultural fear refers to any possibility of an unbalanced acculturation. There have been some concerns that bilingualism could harm students' language heritage and home country cultures. This concern was commonly raised by immigrant parents who came to live in a new country. Some parents worried that their children would adopt a new language together with the new culture and, as consequence, might neglect their heritage culture and heritage language in favour of the new culture or they would lose their language heritage and the culture of their ancestors. The community's apprehension has been argued that it was related to identity (Edwards, 2010): the society want the members of their community to excel in their schools and at the same time they yearn for their language heritage and cultures to be preserved. Their fears were due to the view that the immigrants were expected to replace home country characteristics with the new country's characteristics, in order to make them more like the new country citizens (Dicker, 2003). Regarding this, some schools in these new countries have provided several different bilingual schools that accommodate the needs of the children. It has been suggested that a well-balanced bilingual educational program could minimize this fear (Baker, 2011). The schools ensure that children can

participate well in the classroom, but can also preserve their local language and heritage, at least to some extent.

Educational fear is usually expressed by people who have conceived misguided assumptions about bilingualism and bilingual learning. The vast majority of them represent people who are not directly involved in bilingual research or practices. This expressed fear has less to do with education than with perception about national unity. This fear is concerned more with seeing bilingual education programs as a threat to the state of the majority language (Baetens-Beardsmore, 2003).

The Nature of Bilingual Education

What is Bilingual Education?

Perhaps, before we look into the definition of a bilingual education, it is beneficial to consider what we know about bilingual education and mother tongue development, as has been raised by Cummins (2003), together with some answers, on the following points.

1. Bilingualism has positive effects on children's linguistic and educational development.
2. The level of development of children's mother tongue is a strong predictor of their second language development.
3. Mother tongue promotion in the school helps develop not only the mother tongue but also the children's abilities in majority school languages.
4. Spending instructional time through a minority language in the school does not hurt children's academic development in the majority school language.
5. Children's mother tongues are fragile and easily lost in the early years of school.
6. Negotiation of identity is a crucial factor in minority children's academic success (pp. 61-64).

With the above baseline in bilingual education and mother tongue development, bilingual education has been seen as a controversial and misunderstood matter (Freeman, 2007). Cazden and Snow (1990, p. 9) put it as "a simple label for a complex

phenomenon” and Butler (2013, p. 110) adds “a highly complex social, psychological, and linguistic phenomena”. Its controversial nature is due to the confusion regarding the same term being used for several different programs. The programs could vary in terms of different ideological orientations towards linguistic and cultural diversity, different target populations, and different goals for those target populations (Hornberger, 1991).

In broad terms, bilingual education has been defined as the use of two languages (native and additional language) as the medium of instruction (Brisk, 2005), with the ultimate goal to achieve additive bilingualism, biliteracy and cultural diversity, but the form of the bilingual education can vary greatly (Cummins, 2000). To determine which type of bilingual education is used depends on various variables, and the names of a few are: the native language of the students, which language to take, the language of instruction, and the linguistic goals for the program.

What are the Existing Bilingual Education Models and Practices?

Bilingual education has a diversity of aims, practices and contexts, and where bilingual education has been applied in one country, it has different aims and context to another. For example, bilingual education in the USA or Canada is different from a bilingual context in Malaysia or Indonesia. Bilingual education in the USA or Canada has been associated with minority language students who learn curriculum content in one majority language (for example, English) and in one minority language (for example, Spanish) in the USA context; and for two majority languages, namely French and English in the Canadian context. Bilingual education in Malaysia or Indonesia has been associated with majority language students learning curriculum content in two languages, where one language is a strong ethnic language (namely Bahasa Malay, in the Malaysian context) or a national language (namely Bahasa Indonesia, in the Indonesian context) and the other is English as a second language (in the Malaysian context) or foreign language (in the Indonesian context).

Due to the different aims of bilingual education in different countries and contexts, the perspectives of the different socio-historical contexts need to be considered (Lin & Man, 2009). Ferguson, Houghton, and Wells (1977) provided ten different goals of bilingual education:

1. To assimilate individuals or groups into the mainstream of society; to socialize people for full participation in the community;
2. To unify a multilingual society; to bring unity to multi-ethnic, multi-tribal, or multi-national linguistically diverse states;
3. To enable people to communicate with the outside world;
4. To provide language skills which are marketable, aiding employment and status;
5. To preserve ethnic and religious identity;
6. To reconcile and mediate between different linguistic and political communities;
7. To spread the use of a colonial language, socializing an entire population to a colonial existence;
8. To strengthen elite groups and preserve their privileged position in society;
9. To give equal status in law to languages of unequal status in daily life; and
10. To deepen an understanding of language and culture.

Based on the above ten aims of bilingual education (Baker, 1996) classified bilingual education into four categories:

1. Typical language(s) used by the child in daily life;
2. Typical language(s) in the classroom;
3. The educational and societal aim(s) of the program; and
4. The probable outcomes of the program.

Based on the above ten categories, bilingual education can be classified into three general types and then re-categorised into eight models or programs that have been in practice throughout the world (based on Baker, 2006, 2011). The three general types are: (1) Maintenance programs, (2) Transitional programs, and (3) Enrichment programs.

The specific ten models or programs (summarised from Baker, 2011, pp. 211-250; and (Ovando & Combs, 2012, pp. 35-44) are:

1. Mainstreaming/Submersion education (also known as ‘sink or swim’ method). By ‘submersion’, it metaphorically means that a child is allowed to swim in

the sea of knowledge, without any help. It is believed that the child will quickly learn how to swim in order to survive. In classroom situations, children will try to learn a totally new language without help, and it is hoped that they can cope with that in order to 'survive' in a language classroom. Submersion education is a label used to describe education for a language minority (in the USA context) who are placed in mainstream education with the majority language students. The language of the classroom is the majority language, while that of minority language students is other than English (for example, Spanish or French). Both teachers and students use a majority language (such as English). Thus, the minority language students whose first language is not English are forced to learn English without any help prior to the admission to the classroom. These students need to 'swim' to survive. This type of bilingual education has been practised in countries like the USA.

2. Submersion with withdrawal or pull-out, or sheltered English, or content-based ESL classes. In submersion with withdrawal/pull-out, the language used is the majority language (such as English). The minority language students are assigned to learn the majority language (English) with help under a second language curriculum. These lessons are taught separately from the submersion class. The submersion teacher teaches the academic content and the teacher with the English as a Second Language qualification teaches English. The students are withdrawn or 'pulled-out' from their mainstream class every day for an assigned time for a couple of years. In the submersion with sheltered English, the program contains only minority language students and is conducted in a majority language (in English for non-native students). The teacher uses a simplified form of the majority language and accepts contributions from the students in their native language but does not support the native language of the minority students. In the submersion with content-based English as a Second Language, the teacher teaches both the academic content in the majority language, and the majority language as a second language at the same time. The teachers have both academic content and English as a Second Language qualifications. This type of bilingual education has been practised in countries like the USA.

3. Segregationist education. This education is for minority language students only and is conducted in the minority language, but it is uncommon. This type occurs where minority language speakers are denied access to those programs or schools attended by

majority language students. Such education can be through law and practice. Segregationist bilingual education is aiming at prohibiting the minority language students acquiring the majority language in order that the minority language people do not interfere with the majority language elites. This type of bilingual education has been practised in countries such as South Africa, prior to Nelson Mandela becoming President.

4. Transitional bilingual education (early exit or late exit). This type is also known as Mainstream Bilingual Education or Developmental Bilingual Education. As the term suggests, this type of bilingual education prepares the minority language students with the majority language before they are ready for the mainstream classes. Containing the minority language students only, the students are temporarily allowed to learn academic content in their native language for a maximum of two years. After this period of time, the students join the mainstream classes where the language used is the language of the majority without any language of the minority. For the late exit transitional bilingual education, the student's native language is used in teaching academic content for longer than two years. It places less emphasis on exiting language students from the bilingual program as soon as possible. Students in class receive content area instruction in both languages (40% majority language and 60% native language) until grade 6. Teachers or their assistants need to be bilingual. The aim of this program is majority language monolingualism. This type of bilingual education has been practised in countries like the USA.

5. Mainstream education (with a foreign language taught as a subject). The mainstream education is for students learning academic content using their own home or native language as a majority language and some other foreign language is also learnt. For this class, a foreign language is regarded as a subject (as Maths, Social Sciences, and the like) and the foreign language learnt is not used as a medium of instruction. This type of bilingual education has been practised in Canada, with programs called 'core French' and 'drip-feed'.

6. Separatist education. Separatist education promotes monolingualism in the minority language as well as promoting monoculturalism where possible. This is a way to protect a minority language from being overrun by the language majority for

political, religious, or cultural reasons. This type of bilingual education has been practised in small isolationist religious schools.

7. Immersion bilingual education. The Immersion Bilingual Education is for students to become bilingual and bicultural without the loss of academic achievement. There are four sub-types under Immersion Bilingual Education: Early total immersion, early partial immersion, delayed immersion, and late immersion. These four sub-types of the program relate to the age (early, middle, and later) of the students starting the program, and the amount of time spent in the program (partial or total). This type of bilingual education has been practised in Canada with programs called early total, early partial, delayed immersion, and late immersion. This immersion bilingual education has received some success, due to six important features:

1. Immersion in Canada aims at bilingualism in two majority languages, such as French and English;
2. Immersion bilingual education is not compulsory but optional;
3. Home language is allowed for one to one and half years as the classroom language of instruction;
4. Teachers are bilinguals with both French and English qualifications;
5. The aim of classroom language communication is to be meaningful, authentic and relevant to children's needs;
6. Most students are monolingual when starting the immersion program. They start the program with a similar lack of experience of the second language. Therefore, the students have high self-esteem and motivation because no other students possess a higher level of language skills.

8 Maintenance and heritage language bilingual education. This education program occurs where the language minority children use their native, ethnic, home, or heritage language as a medium of instruction with the goal of full bilingualism. The six main features of this type of bilingual education are as follows:

1. Most of the time, students will come from a minority language home, but at the same time the minority language may be the majority language of the local community (for example, Spanish language is used in Spanish communities in some cities in the US). The Heritage Language Education

program teaches students through a heritage language, and includes not only heritage language students but also other native students (whose first language is English);

2. Parents have the choice to send their children to either mainstream schools or to heritage bilingual schools;
3. The language of minority students is used for half or more of the curriculum time;
4. Concepts are not retaught in the other language. Transfer in the two languages is encouraged;
5. The justification given to the heritage languages is that the minority language is easily lost, while the majority language is easily maintained. Thus, bilingualism is obtained from the initial concentration on the minority language at school. In the later stages of elementary school, attention is focused more on majority language development;
6. Heritage language schools are mostly elementary schools. However in other places, the heritage schools start at kindergarten (in Hawai'i) or at the end of secondary school (in Wales).

In the USA, this type of program is also called Maintenance Bilingual Education, or Developmental Maintenance Bilingual Education. Maintenance and Heritage Language Bilingual Education has been practised in countries like the US for Navajo and Spanish languages, Canada for the Ukrainian language, and New Zealand for the Maori language.

9. Dual Language Immersion. The aim of this type of bilingual education is language separation and compartmentalisation. There is only one language used in each period of instruction. Language boundaries are established in terms of time, curriculum content and teaching. Regarding time, a decision is made about when to teach through language on alternative days, or different lessons may use different languages. In regards to the curriculum content, some lessons are taught using one language and the other lessons are taught in another language. With regards to teaching, teachers are not allowed to switch languages when teaching in a particular language.

In the USA, Dual Language or Bilingual Education has the following features:

1. A non-English language (that is a minority language) is used for at least 50% of instruction that lasts for up to six years;
2. In each period of instruction, only one language is normally used. Instruction must be adjusted to the students' language level, but must be challenging, empowering and enabling. Language is learnt primarily through content;
3. Both English and non-English speakers are present in approximately balanced numbers and integrated for most content instruction. The English and non-English speakers are integrated in all lessons.

Having a mission to produce bilingual, biliterate and multicultural children, this type of bilingual education has five major goals:

1. High levels of proficiency in students' first language and a second language;
2. Reading and writing at grade level in both languages;
3. Academic achievement at, or above level (e.g. mathematics, science, social studies);
4. Positive intercultural (multicultural) attitudes and behaviours; and
5. Communities and society to benefit from having citizens who are bilingual and biliterate, who are positive towards people with different cultural backgrounds, and who can meet national needs for language competence and a more peaceful coexistence with peoples of other nations.

In order to meet the mission, a variety of practices are utilized, such as the following:

1. The two languages of the school are given equal status in the school;
2. The school ethos is bilingual;
3. In some Dual Language bilingual schools, the two languages are taught as languages (sometimes called language arts instruction). In some other Dual Language bilingual schools, the two languages are used as the medium of instruction in order to ensure bilingual development;
4. Staff in dual language classrooms are often bilingual; and
5. The length of the Dual Language bilingual program needs to be longer rather than shorter. Such a program for two or three grades is insufficient.

This type of bilingual education, also known as Two-Way Schools, Two-Way Immersion, Two-Way Bilingual Education, Dual Language Education, Bilingual Immersion, Spanish Immersion, Double Immersion, and Interlocking Education, has been practised in some countries like the USA (mainly elementary schools).

10. Bilingual education in majority languages

Another type of bilingual education program refers to learning a second language by students living in Asian and other countries who are originally already bilingual or multi-lingual (see Table 2.2). Thus, a second language is added as a medium of instruction to the original strong indigenous or national language. Therefore this means that some curriculum content is learnt through students' second language. This type of bilingual education has been practised in countries like Luxembourg and some European Schools.

Baker (2011) summarised the types of bilingual education that have been presented earlier. Types of bilingual education are presented in Table 2.2.

Table 2.2
Type of Bilingual Education

MONOLINGUAL FORMS OF EDUCATION FOR BILINGUALS				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
MAINSTREAMING/ SUBMERSION (Structured Immersion)	Language Minority	Majority Language	Assimilation/ Subtractive	Monolingualism
MAINSTREAMING/ SUBMERSION with Withdrawal Classes/ Sheltered English/Content-based ESL	Language Minority	Majority Language with 'Pull-out' L2 lessons	Assimilation/ Subtractive	Monolingualism
SEGREGATIONIST	Language Minority	Minority Language	Apartheid	Monolingualism
WEAK FORMS OF BILINGUAL EDUCATION FOR BILINGUALS				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
TRANSITIONAL	Language Minority	Moves from minority to majority language	Assimilation/ Subtractive	Relative Monolingualism
MAINSTREAM with Foreign Language Teaching	Language Minority	Majority language with L2/FL lessons	Limited Enrichment	Limited Bilingualism
SEPARATIST	Language Minority	Minority Language (out	Detachment/ Autonomy	Limited Bilingualism

of choice)

STRONG FORMS OF BILINGUAL EDUCATION FOR BILINGUALISM AND BILITERACY				
Type of Program	Typical Type of Child	Language of the Classroom	Societal and Educational Aim	Aim in Language Outcome
IMMERSION	Language Majority	Bilingual with initial emphasis on L2	Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
MAINTENANCE/HERITAGE LANGUAGE	Language Minority	Bilingual with emphasis on L1	Maintenance, Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
TWO WAY/ DUAL LANGUAGE	Mixed Language Minority & Majority	Minority and Majority	Maintenance, Pluralism and Enrichment. Additive	Bilingualism & Biliteracy
MAINSTREAM BILINGUAL	Language Majority	Two Majority Languages Pluralism	Maintenance, & Biliteracy and Enrichment. Additive	Bilingualism

Note: L1=First Language; L2=Second Language; FL=Foreign Language

Source: (Baker, 2011, pp. 209-210; and Ovando & Combs, 2012, pp. 35-44)

Considering the ten models of Bilingual Education above, this study would suggest that bilingual education in Aceh Province falls into the last model which is model number 10, due to the fact that the Acehnese type of bilingual education program refers to learning a foreign language by students living in Aceh Province of Indonesia when they have known a local language, such as Acehnese, Bahasa Jamee, Bahasa Tamiang and another nine local languages, and they have known Bahasa Indonesian as the national language. Thus, a foreign language is added as a medium of instruction to the national language. Therefore this means that some curriculum content is learnt through students' foreign language, in this case English.

Brief History of Bilingual Education Worldwide

Historically, bilingual education in the world has existed for as long as a thousand years (Mackey, 1978) when human society was formed and then continued through from the ancient world to the renaissance and to today's modern world (Lewis, 1977, cited in Baker 2011). Bilingual education in the world, for example, in the USA, Sweden, England, and Canada, has been associated with each country's historical context of immigration, as well as political movement (Baker, 2011). Bilingual education for some other countries like Wales and Ireland has been linked with the movements in language rights and self-government (Jones & Martin-Jones, 2004). In

Japan, it has been connected to a movement from monolingual ideology to internationalism (Maher, 1997). In Southeast Asia (such as in Indonesia, Malaysia, Brunei Darussalam, and Singapore), bilingual education has been related to the birth of ASEAN and English as a *lingua franca* in ASEAN (Kirkpatrick, 2012; Kirkpatrick & Sussex, 2012).

Well-known contexts for bilingual education include the USA and Canada. Bilingual education in the USA was believed to have started when immigrants from South America, Italy, Germany, Holland, France, Poland, Czechoslovakia, Ireland, Wales, Spain, China, and other countries, came to the USA (Baker 2011). Putting faith in the country's promise of liberty and a better life, the immigrants entered the country together with their heritage, languages and cultures. It did not take long before the immigrants were assimilated and schools for their children were set up, and it appeared that a competition between public and private schools that later motivated an openness for the language of the immigrants occurred. However, bilingual education in the USA received a setback when the country prohibited the teaching of certain languages due to an anti-German sentiment, related to the First World War. Bilingual education in the country resumed in 1963, when the first modern dual school was established in South Florida (Baker 2006). Since that time, bilingual education has become considerably accepted after a lawsuit to have it formally implemented.

The following is a summarised history of bilingual education in the USA based on work by Baker (2006, 2011).

Table 2.3
Brief History of Bilingual Education in the USA

Year	US Legislation/Litigation affecting Bilingual Education	Implication
1906	Nationality Act passed	First legislation requiring immigrants to speak English to become naturalized.
1923	<i>Meyer v. Nebraska</i> ruling the US Supreme Court	The ruling outlawed, as an unconstitutional infringement of individual liberties, arbitrary restrictions on teaching languages other than English. Proficiency in a foreign language was also constitutional.
1950	Amendments to the Nationality Act	English literacy required for naturalization.
1954	<i>Brown v. Board of Education</i>	Segregated education based on race made unconstitutional.
1958	National Defense Education Act	The first federal legislation to promote foreign language learning.
1965	Immigration and National Act	The Act eliminated racial criteria for admission,

		expanding immigration especially from Asia and Latin America. The Act also emphasized the goal of ‘family unification’ over occupational skills. This encouraged increased immigration by Mexicans in particular.
1965	Elementary and Secondary Education Act (ESEA)	Funds granted to meet the needs of ‘educationally deprived children’.
1968	Elementary and Secondary Education Act (ESEA) amendment: The Bilingual Education Act, Title VII	Provided funding to establish bilingual programs for students who did not speak English and who were economically poor.
1974	<i>Lau v. Nichols</i>	Established that language programs for language minorities not proficient in English were necessary to provide equal educational opportunities.
1974	Equal Educational Opportunity Act (EEOA)	Codified the <i>Lau v. Nichols</i> decision, requiring every school district to take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs.
1974	Reauthorization of Bilingual Education Act Title VII of ESEA	Native-language instruction was required for the first time as a condition for receiving bilingual education grants. Bilingual Education was refined as transitional (TBE).
1975	Lau Remedies	Informal guideline on schools’ obligations toward LEP students. This required the provision of bilingual education in districts where the civil rights of such students had been violated.
1976	<i>Keyes v. School District no. 1, Denver, Colorado</i>	Established bilingual education as compatible with desegregation.
1978	Reauthorization of Bilingual Education Act Title VII of ESEA	A new restriction was introduced. Grants could support native-language instruction only to the extent necessary to allow a child to achieve competence in the English language. Funding was thus restricted to TBE; maintenance programs were now ineligible for funding. The term ‘Limited English Proficient’ (LEP) introduced, replacing LES (Limited English Speaking).
1980-81	Lau Regulations	The Carter Administration attempted to formalize the Lau Remedies, requiring bilingual instruction for LEP students where feasible. The Reagan Administration subsequently withdrew the proposal, leaving uncertainty about schools’ obligations on this area.
1981	<i>Castañeda v. Pickard</i>	An Appeal court decision established a three-part test to determine whether schools were taking “appropriate action” under the 1974 Equal Educational Opportunity Act. Programs for LEP students (bilingual or otherwise) must be: (1) based on sound educational theory, (2) implemented with adequate resources, and (3) evaluated and proven effective.
1983	US English Movement launched	Debates about the dominant place of English in law, society and education became more prominent.
1984	Reauthorization of Bilingual Education Act Title VII of ESEA	While most funding was reserved for TBE, monies for maintenance programs were once again permitted, along with ‘special alternative’ English-only programs.
1988	Reauthorization of Bilingual Education Act Title VII of ESEA	Same as in 1984, but 25% of funding given for English-only Special Alternative Instructional (SAIP) programs.
1994	Reauthorization of Bilingual Education Act Title VII of ESEA	Full bilingual proficiency recognized as a lawful educational goal. Funded dual language programs that included English speakers and programs to support Native American languages. The quota for funding SAIP programs was lifted. The new law sought to bring LEP students into mainstream school reform efforts, making it

		more difficult for their particular needs to be ignored in policymaking.
1998	Proposition 227 passed in California	The 'Unz initiative' sought to impose severe restrictions on native-language instruction for English learners in California. Most bilingual programs dismantled, with similar measures in Arizona (2000) and Massachusetts (2002).
2002	No Child Left Behind legislation as a reauthorization of the Elementary and Secondary Act of 1965 and a repeal of the Bilingual Education Act	Schools and states encouraged to move to English-only education through mandatory high-stakes testing in English. Measures of Adequate Yearly Progress (AYP) reported for schools, school districts and states including English proficiency.

Source: Baker (2006, pp. 201-202; 2011, pp. 196-198)

With immigrants accounting for two thirds of the population, Canada is one of the most culturally and linguistically diverse countries in the world, representing a bilingual and pluralistic country (Clark, 2012). Canada has been famous for its immersion programs. The Canadian immersion program started in Quebec when it underwent a social transformation in the 1960s. Some English speakers went out of the city while others stayed and learnt French to compete in the markets. Originally, the St Lambert's immersion program did not contain the idea of bilingualism, and consisted of four aims (Roy & Galiev, 2011):

1. To allow students to be competent both in oral and written French;
2. To maintain the development of English as the first language;
3. To allow students to learn content according to their age and level of school; and
4. To aid French speaking students and English speaking students to develop understanding and respect for each other's culture and language while preserving their own culture and language (summarised from Roy & Galiev, 2011).

The success of the St. Lambert's immersion education program promoted bilingualism which was then used to transfer to other schools where English was dominant.

In Southeast Asia, English has been spread remarkably well. Its spread has been associated with the establishment of the Association of Southeast Asia Nations (ASEAN). Representing ten countries: Indonesia, Malaysia, Singapore, the Philippines, Thailand, Brunei Darussalam, Vietnam, Laos, Myanmar, and Cambodia, it was founded in 1967 to stand against communism, as communism was then seen as a threat in the region. Even though there was no agreement to use English as the language for communication among the ASEAN countries, English was chosen as the *de facto* language for communication, also known as a *lingua franca*, and the decision was considered 'natural' (Kirkpatrick, 2007, p. 137). In 1999, it became the official language for ASEAN (Kirkpatrick, 2012).

English has played an important role in the curriculum of ASEAN countries. It has been taught in schools with varying degrees and emphasis. Some countries have regarded it as a medium of instruction, and some other countries have used it as a core subject. English is the medium of instruction for all subjects in Singapore, as it is in the Philippines and Brunei Darussalam. In Malaysia, it has been the medium of instruction for 11 years (from 2002 to 2013). In Indonesia, English has been taught as a foreign language for a number of years, and it is also regarded as a medium of instruction for some levels and types of schools. In Myanmar, English is taught as a separate subject (Kirkpatrick, 2012).

In recent years, however, there has been some controversy in some ASEAN countries as to whether English is best regarded as a core subject only, or as a medium of instruction. Different countries have dealt with this controversy differently. Malaysia, which has used English as a medium of instruction for subjects like Mathematics and Science in secondary schools since 2002, reversed this decision in 2013 and returned to Bahasa Malaysia as the medium of instruction. The reversal was due to complaints that the use of English as the medium of instruction only benefitted the urban area-students, and the rural area-students were left behind (Kirkpatrick & Sussex, 2012). Similarly, Indonesia has in 2013 changed the use of English, from its position as a medium of instruction for some types and levels of schools, to teaching it only as core subject with a lesser amount of class time. The reason behind this decision was based on certain developing views that Bahasa Indonesia should be emphasized and regarded as the medium of instruction. In contrast to Malaysia, Brunei Darussalam has increased the

amount of English used as the medium of instruction, leading to its use as the medium of instruction for Primary 1, for certain parts of the curriculum (see Kirkpatrick & Sussex, 2012).

The Nature of Bilingual Education in Indonesia

What is the Position of Bahasa Indonesia?

Bahasa Indonesia (also called Indonesian) has been regarded as the Indonesian national language since *Sumpah Pemuda* declaration (The Youth Pledge) on October 28, 1928. The pledge contained three oaths, that is: (1) to acknowledge one's motherland: that is Indonesia; (2) to acknowledge one nation: that is Indonesia; and (3) to uphold one language of unity: Indonesian. Bahasa Indonesia was utilised as the national language well before the declaration of Dependence Day took place in August 17, 1945. Bahasa Indonesia has been regarded as the most important language in Indonesia even though it was not the most spoken language. The reason for choosing it as the Language of Unity was because it had been the least 'treated' language compared to Javanese which had been spoken widely in Indonesia. At that time, Javanese was a language of elite politics in Indonesia. Should Javanese be taken as a national language, it would lead to a conflict of interest amongst the Indonesians, since it would bring privilege to certain powerful community groups and it would emphasise the role of the hierarchy that culture in the language brought with it (Kirkpatrick, 2011).

As a language that unites 726 dialects spoken by more than 400 ethnic groups in Indonesia, Bahasa Indonesia has been a great success (Kirkpatrick & Sussex, 2012). Today, it is spoken by more than 200 million Indonesians (Hamied, 2011, in Kirkpatrick, 2011), and it functions well in administration, politics, and the judiciary. Bahasa Indonesia is now the main medium of instruction in educational institutions at all levels throughout the country.

Why is English in the Indonesian Curriculum?

English has been taught as a foreign language in Indonesia as a compulsory subject at secondary levels. For elementary levels, it was not compulsory, and therefore was not part of the English curriculum. However, some elementary schools have taught it as part of their local contents, due to school preferences and prestige. As local content,

it was not part of the curriculum assessment, so it was not tested for the end-of-academic-year exit examinations.

The paradigm in the Indonesian educational context was changed after 2003 in regards to a new perception towards the role of English as an international language and as the language of technology. The government of Indonesia, on behalf of the Ministry of Education, believed that the country would benefit by making English compulsory for Indonesian students at secondary level, so that students could master English as a foreign language. This decision was implemented in 2003 by the Act of the Indonesian Republic Number 20, Article 50, and Verse 3. The Act advocated that it was imperative for the Indonesia government and for local governments to establish at least one unit of English education with the international standard at all levels of education. The aim of this Act was to ensure that Indonesian citizens master English skills in order to be able to compete in different fields, such as education, information technology, communication, trade, and social culture.

Why Do They Have Bilingual Education in Indonesia?

The Act has been translated as the provision of bilingual programs in certain schools nation-wide. A number of schools were established as pilot schools, named as “International Standard School” (SBI or *Sekolah Bertaraf International*) and “Prospective International Standard School” (RSBI or *Rintisan Sekolah Bertaraf International*), while regular schools functioned normally. These schools replaced the use of Bahasa Indonesia with English. The criteria for an International Standard School are as follows (Depdiknas, 2007):

1. English is to be used as the medium of instruction for science, mathematics and core vocational subjects from year 4 of primary school and throughout junior secondary school, senior secondary school and vocational secondary school;
2. Teachers must possess the competence required to teach their subjects through English; and
3. Head teachers must possess active mastery of English (Depdiknas, 2007).

In order to be appointed as a Prospective International Standard School, certain “criteria need to be met among which is the teaching of at least two subjects in English”

(Chodidjah, 2009, p. 88). In implementing bilingual education, the Indonesian government and local governments have facilitated some innovative programs, such as ensuring bilingual teachers are actively involved with in-service training and sending the teachers to study abroad. Such programs were intended to help bilingual teachers equip themselves with the knowledge and skills needed for success in teaching in the bilingual program. In addition to that, the government also supported the schools with modern facilities in order to enable the teaching and learning bilingual programs to run as expected.

The bilingual program gained success after a couple of years of implementation. However, along the way, it also received many criticisms. Opponents of bilingual education have criticised the government for not being ready to carry out the bilingual program to the standard required in terms of school facilities, teachers, and students. While the idea of bilingual education was positively accepted, the implementation of it has been seen as immature, in term of insufficient preparation or understanding of what was needed (Davies, 2005). Bilingual schools have also been associated with luxury and prestige, due to the fact that only students from the richer families were able to enrol in them. Students from poor families could not afford it, since some bilingual schools required financial support from students (Revianur, 2013b). Another criticism focused disagreement about Bahasa Indonesia being replaced by English. A fear was raised that English would diminish students' valuing of the Indonesian language and culture (Tubagus, 2005).

Bilingual Education Program Ban in Indonesia (2013)

Due to criticism that bilingual education was seen as discriminating against a fair education, some groups brought this case to the Judicial Court (known in Indonesia as *Mahkamah Konstitusi*). With regards to this issue, the Judicial Court came to the agreement that bilingual programs had to be eliminated from the Indonesian education system because they violated the basic law of the Indonesian Republic (Year 1945) and they did not have a strong supporting background case (Revianur, 2013b). The previously bilingual schools returned to standard programs (Revianur, 2013a). This ban on bilingual education programs came into effect in January 2013.

Coinciding with the banning of bilingual education in Indonesia, a new 2013 curriculum has been launched. The primary change regarding English stated in the new curriculum was that there would be no bilingual education in the Indonesian educational system and the amount of teaching English has been drastically decreased. This change in regards to the English curriculum has a number of effects. It has been predicted that many English teachers would lose their jobs due to the omission (or marked decrease) of their teaching hours. Students who have previously benefited through bilingual educational programs may find it difficult, or unhelpful, to return to the original mode of separate English teaching classes. It would be expected to take some time for many previously-taught bilingual students to readjust. Handbooks and selected texts and curriculum books would need to be replaced with those re-written for the new curriculum and this may take some time also.

What is the future of Bilingual Education in Indonesia?

It would seem that it is not easy to see much of a future for bilingual education in Indonesia in the near future from 2013 onwards, especially with the implementation of the new curriculum where bilingual education is banned by law. However, as the trends in the Indonesian educational system are dynamic, one can hope that a revised decision could be made that benefits Indonesian students in terms of having opportunities to learn additional languages in order to enrich and prepare students for a global village of bilingualism and multilingualism.

A Review of Bilingualism and English as a Second or Foreign Language

There have been a reasonable number of studies conducted on second language acquisition, bilingualism, English as a second language, and English as a foreign language, in various contexts with different data analyses and measures. This section revisits three areas of the research that focused on: (1) Review of second language acquisition, (2) Findings on experimental and control groups relating to reading and writing, and (3) Findings on Rasch measures in second language.

Review of Second Language Acquisition

A review of second language acquisition has recently been conducted by Dixon et al. (2012) and published in the Review of Educational Research Journal. The study investigated relevant papers from 72 peer-reviewed journals that had been published from 1997 to 2011. Due to its concise and complete work, the study (pp. 5-60) outcomes are summarised for this section.

The study entitled “What we know about second language acquisition: A synthesis From Four Perspectives”. It looked at four perspectives on second language acquisition, namely; (1) Foreign language educators; (2) Child language; (3) Sociocultural; and (4) Psycholinguistics. Each perspective focuses on a different unit of analysis, context, age of learners, language domain focused on, and major questions (see Table 2.4).

Table 2.4
Difference among the Four Perspectives

Perspective	Unit of analysis	Context	Age of learners	Language domain focused on	Major questions
Foreign language (FL) educators	Student achievement or proficiency	FL classroom with little FL exposure in community	Adolescents and adults	Accent, grammatical correctness, reading ability	Developing and improving teaching methods and techniques
Child language	Child	Naturalistic, extended to teacher or student in classroom	Young children	Grammar, lexicon, extended discourse (narrative, academic language)	Identifying factors that influence child L1 development, bilingual L1 acquisition, child L2 acquisition
Sociocultural	Students and interlocutors	Naturalistic or classroom	Any	Communicative effectiveness, pragmatics	Understanding social and cultural differences, impact of identity, interpersonal interactions
Psycholinguistics	Cognitive and brain processes	Laboratory, extended to classrooms, naturalistic	Any	Grammar, lexicon, text structures	Understanding process of L2 acquisition, transfer of L1 skills to L2.

Note: L1= First language; L2=Second language

Source: Dixon, Zhao, Shin Wu, Burgess-Brigham, Gezer, & Snow, 2013, p. 8.

The review was carried out to address five questions of particular relevance to the four perspectives and their relevance to second language education. They are:

1. What are the optimal conditions for second language acquisition?
2. What are the characteristics of excellent or unsuccessful second language learners?
3. What are the characteristics of excellent or unsuccessful second language teachers?
4. What are reasonable expectations for speed and accomplishment for second language learners of different ages?
6. Has information generated by the four research perspectives influenced the formulation of educational policies for second language learners?

The findings provided in the review were based on answers to the five questions above.

Findings for Question 1: What are the optimal conditions for second language acquisition?

There was no confirmed definition on ‘optimal conditions’ for the best way to teach second language learners. Optimal conditions for acquiring a second language for different populations vary in regard to learning contexts, pedagogical goals, program setup, learner characteristics, and the interactions among these variables. The best conditions for second language learners in second language-majority contexts include higher family socioeconomic status together with parent and grandparent education, strong home literacy practices, opportunities for informal second language use, and well-designed and well implemented educational programs specifically for second language literacy instruction. Regarding this, teachers can make a positive difference by encouraging home literacy, sending home books and other literacy materials and getting the parents to read with their children in either second language or first language. Further positive help can be provided by taking them to the library, and supporting informal second language use by mixing second language learners with first language speakers, as well as promoting integrated curricular activities and ensuring that there is

sufficient time to implement educational programs and lesson plans that follow research-tested designs in their schools

Optimal conditions for second language learners in a foreign language setting were stated as explicit instruction helps students, especially in learning grammar, and using academic content to teach the second language may be beneficial to building vocabulary in a second language. However, it was also stated that the intensity of second language instruction seemed to make little or no difference in second language learning performance.

Findings of Question 2: What are the characteristics of excellent or unsuccessful second language learners?

The following findings summarise the answers to Question 2. A positive aptitude and motivation of second language learners contribute to the largest positive difference in second language outcomes. Better first language skills, lower second language anxiety, and gender (females do better at language) contribute to positive second language outcomes. Verbal ability is more important to older second language learners than to younger second language learners. First language literacy skills are reasonable predictors of second language literacy skills and the outcomes for foreign language students and for learners in second language-majority settings are influenced by different factors.

Findings of Question 3: What are the characteristic of excellent or unsuccessful second language teachers?

Competent second language teachers possess adequate proficiency in the target language. They desire to teach well and there is a positive correlation with self-efficacy, intellectual excitement, and teacher reflectivity. Good teachers have good classroom organization where instruction is clear and well planned, and students know what to expect. Good second language teachers have, at least, some of the students' first language proficiency and they know when and how to use it.

Findings of Question 4: What are the reasonable expectations for speed and accomplishment for second language learners of different ages?

For learners in second language-majority contexts, young second language learners are likely to be successful at second language only after several years. These second language learners can reach native-like proficiency in oral fluency, vocabulary, and literacy tasks, even when they start after the age 12, but for the latter, this proficiency does not necessarily include grammar and pronunciation. In the USA context, second language learners who arrived as teenagers succeeded academically in the second language, given appropriate input, instruction, and motivation.

Second language learners' first language skill affects their rate of second language acquisition. Learners, whose first language was more distant to the second language (such as Korean and English), took longer to acquire competence. For second language learners in foreign language contexts, the canard that says 'younger is the better' should be rejected and, holding hours of instruction constant, older learners performed better on the proficiency measures but, of course, an early start could be beneficial for providing more hours of input.

Findings of Question 5: Has information generated by the four research perspectives influenced the formulation of educational policies for second language learners?

In the USA, the Act of No Child Left Behind (NCLB) had unforeseen consequences for second language learners and second language education. The Act was not supported by the four perspectives. USA state policies indicated that one year was considered sufficient for second language learners to gain English proficiency in a mainstream classroom. This, however, was not supported by the research from the four perspectives. Two-way bilingual education was effective at the pre-school level and some two-way programs with voluntary enrolment were effective at the pre-kindergarten to 8 levels.

There was no national policy (in the No Child Left Behind Act) in regard to foreign language teaching or second language learning. USA policy did not incorporate the research findings of any of the four bodies of research work on English in the second language learner context or in any other language.

(Summarised by Khairiah Syhabuddin from Dixon, Zhao, Shin Wu, Burgess-Brigham, Gezer, & Snow, 2013, pp. 5-60, from the Review of Educational Research Journal)

Findings on Experimental and Control Studies for Reading and Writing, and Attitude and Behaviour

There have been a number of studies using experimental and control groups on the four sub-skills of English proficiency, namely, listening, speaking, reading, and writing. This section, however, only revisits studies on reading and writing due to the fact that they were the core interest of the present study. The studies on reading and writing are revisited prior to the studies on attitude and behaviour.

There have been a number of recent experimental and control group studies related to reading comprehension. For example, Alshumaimeri and Almasri (2012) conducted an experimental and control study on the effect of using WebQuests on Reading Comprehension performance. The respondents for the experimental group were 42 male Arab students and they received traditional teaching together with WebQuest additional activities. The respondents for the control group were 41 male Arab students and they received only traditional teaching. The performance data were analyzed using SPSS and t-tests. The results indicated that the WebQuests had potential in promoting reading comprehension.

Buyuktaskapu (2012) carried out a study to examine the effect of family support during pre-reading training on reading success in a Turkish primary school. There were 25 first graders participating in the Family Supported Pre-Reading and they were taken as the experimental group. The control group consisted of 25 first graders who attended nursery class. Using various true-score non-linear scales of reading and writing to analyze the data, the study revealed that children attending the Family Supported Pre-Reading Program had more reading success than children attending nursery classes.

Nakanishi and Ueda (2011) explored the effect of extensive reading and shadowing on reading comprehension. Shadowing has been defined as a task of listening in which students track the target speech and repeat it as immediately and quickly as possible without referring to a text. The study used 89 first-year Japanese university students to comprise two experimental and one control groups. Data were collected using the Secondary Level English Proficiency Test of students' reading comprehension and analyzed with ANOVA to compare groups. The results showed that there was no statistically significant dissimilarity among groups.

Sahin (2011) analyzed the effect of the way Turkish students read a reading comprehension text, that is, through page scrolling or page-by-page reading, in relation to the students' reading comprehension performance. Forty-six fourth grade students who were in one experimental group and one control group were the participants of the study. T-tests were used to analyze the data. The result suggested that there were no significant differences between the comprehension scores of students using the scrolling reading method and the scores of students using the page-by-page reading method.

Bolukbas, Keskin, and Polat (2011) conducted a study on the effectiveness of cooperative learning on reading comprehension skills. The respondents were 40 Turkish university level students, consisting of 20 students who used a cooperative learning technique as the experimental group and 20 students taught using a traditional teaching model. SPSS 11.00 was used to analyze the data. The results confirmed that cooperative learning is more effective in improving learners' reading comprehension skills.

Novel studies on writing have also been conducted. Tajeddin and Daraee (2013) studied the effect of form-focused and non-form focused tasks on vocabulary learning through written forms of English as a foreign language. This study used 50 Iranian students who learnt English as a foreign language, consisting of two experimental groups and one control group. Data were analyzed using ANOVA and resulted in various conclusions, indicating that task effectiveness depended upon the level of involvement and the task's influence and complexity.

Tabatabaei and Assefi (2012) explored the effects of a portfolio assessment technique as a teaching, learning, and evaluation instrument on EFL learners' writing performance. Forty Iranian students majoring in English participated in the study and were divided into an experimental and a control group. The portfolio assessment was assigned to the experimental group and the traditional approach of writing assessments was assigned to the control group. Using the TOEFL and IELTS tests, the result indicated that the students with portfolio assessment outperformed their counterparts in both writing performance and writing sub-skills.

Hosseini (2012) conducted a study to investigate the effectiveness of asynchronous computer-mediated corrective feedback, both explicit and implicit, on increasing the prepositions' correct use. The respondents were 45 Iranian elementary

EFL learners, who were assigned to two experimental groups and one control group. The study used a one-way ANOVA and concluded that the students who received explicit corrective feedback outperformed the students who only received implicit corrective feedback and those who received no corrective feedback.

Marashi and Dadari (2012) explored the impact of task-based writing on Iranians' learning ESL and their writing performance and creativity. The participants were 56 students grouped into experimental and control groups. Using the Abedi-Schumacher Creativity Test and Cambridge General Mark Scheme for Writing, the results showed that learners benefited significantly from task-based writing and creativity.

There has been a reasonable number of studies conducted on attitude and behaviour with regard to learning English as a second or foreign language (see Alarcon, 2011; Ali, Mukundan, Baki, & Ayub, 2012; Gebhard, 2012). However, only a few studies have been conducted using experimental and control groups. Soleimani, Mainnzadeh, Kassaian, and Ketabi (2012) aimed their study at the effect of instruction based on Multiple Intelligence Theory with regard to attitude and learning of general English. The study recruited 61 Iranian female and male students who were randomly divided into one experimental class and one control class. ANCOVA and independent sample t-tests were used to analyze the data. The results indicated that students taught on the basis of Multiple Intelligence Theory exceeded students who were taught traditionally. They outperformed both in general and sub-skills of learning English. The results also revealed that students who were taught on the basis of Multiple Intelligence Theory significantly improved their attitude toward learning English.

Mekheimer (2012) conducted research on the effect of using Blackboard technology (e-learning) and online dictionaries with English as a foreign language, based on students' translating skill and attitude towards learning English. The study involved 83 male Saudi Arabian university students who were assigned to experimental and control groups. The data were analyzed with t-tests and showed students' translation skills improved together with improvements in positive attitudes toward translation.

Merisuo-Storm (2007) investigated students' attitudes, with regard to foreign language learning and the development of literacy skills in bilingual education. The research consisted of two studies carried out within two years. The first study used 78 students who were divided into three experimental groups, and 58 respondents who joined three control groups. The second study involved 70 students in three experimental groups and 75 students in three control groups. Data were analyzed using the SPSS program and the results showed that students' literacy skills in the bilingual classes were significantly better than those in the monolingual classes. In addition to that, the results indicated that the students in bilingual classes showed significantly more positive attitudes towards foreign language learning than the students in the monolingual classes.

Yoshimura (2006) carried out a study on whether manipulating foreknowledge (schemata) of output tasks leads to differences in students' reading behaviour, text comprehension, and noticing of long forms. Participants for the study were 57 Japanese university students who sat for three different experimental tests. The data were analysed using ANOVA and the results suggested that foreknowledge of output tasks influenced the students' behaviour. It also revealed that the students performed more translation into their first language, expressed more interrelation between their target language and inter-language, and engaged more with language form when foreknowledge manipulation of output tasks was used.

Despite the scarcity of gender studies in EFL research, in terms of pretest-posttest studies, there are studies focusing on gender in relation to students' reading comprehension and writing achievement, and attitude and behaviour. Rahmani and Sadeghi (2011) conducted a study to examine the process and product effect of note-taking strategy on reading comprehension and written material, with gender as a moderating variable. The respondents were 108 undergraduate Iranian EFL students who were assigned into two groups. The experimental group received training on how to take notes and to use graphic organisers as guide. The control group did not receive any instruction. The results of the Two Way ANOVA suggested that the experimental group performed significantly better, remembered more important ideas, and better identified the relationship between ideas. There was no statistically significant effect on gender on students' performance in the comprehension and retention tests. A study to

investigate the effect of multiple intelligence strategies, covering logical-mathematical intelligence and interpersonal intelligence and interpersonal intelligence was conducted by Abdelrahman and Jallad (2008). The study took ninth graders for its respondents who comprised of two male sections and two female sections. Using the Two Way ANOVA, the study showed that there was a significant difference in the students' reading comprehension due to the teaching strategies in favour of the experimental group. There was no significant difference in the students' reading comprehension in regards to the students' gender.

Jafari and Ansari (2012) conducted a research on students' writing achievement. They investigated the effect of group work on Iranian EFL learners' writing accuracy and the effect of gender on text production. The respondents were divided into one experimental group and one control group. The experimental group wrote collaboratively while the control group wrote individually on four essay writing sessions with the same topics and genre. The results revealed that the experimental group outperformed the control group, and females outperformed males. Sadeghi and Sharifi (2013) investigated the effect of post-teaching activity, comprising game, narrative writing, role-play, and speaking tasks on vocabulary gain of EFL learners. The sample of study was 111 elementary female and male adult EFL learners. The respondents were grouped into four experimental groups for females and four experimental groups for males as well as two control groups for each gender. The results of the Two-Way ANOVA indicated that there were statistically significant main effects for vocabulary learning across different activities that lead to the higher vocabulary gain. Female learners outperformed male learners.

Murad Sani and Zain (2011) conducted a study on attitude and behaviour. They investigated the relationships among second language reading attitudes, reading self-efficacy, and reading ability, as well as gender differences in a non-supportive ESL setting. The respondents were 218 students who completed a translated version of Middle/Secondary Reading Attitude Survey and English reading measure. The result stated that the students' second language reading ability was average, their second language reading self-efficacy was low, and the respondents' attitudes were not positive. There was no significant gender difference in self-efficacy. However, female students showed more promising attitudes and comprehended significantly better.

Arshad and Ali (2013) conducted a study to explore whether gender differences effect on learning English in a culture of Pakistan. A self-reporting questionnaire based on a 5-point Likert Scale was used to collect the main data. The participants were students of 'immediate class' and the data were analysed with SPSS. This suggested that social factors have a great effect and impact on learning the English language.

Findings Using Rasch Measures and Control and Experimental Groups

While some research has been done with Rasch-created linear measures on reading, writing, and attitude and behaviour (see Knoch, 2010; Koh, 2008; Lim, 2011; Metsämuuronen, Svedlin, & Ilic, 2012; Wang, Kim, Bong, & Ahn, 2012), research with Rasch measures conducted using experimental and control groups for English as a second language remains scarce. Waugh, Bowering, and Torok (2005) created linear Rasch scales to measure reading comprehension, and attitude and behaviour for Thai English as a second language students. The study explored the use of various genres to improve the teaching and learning of English reading. The participants were 300 grade 7 students who were taught English as a second language through genre-based methods, which were called Expository Genre, Narrative Genre, and Journalistic Genre, and were compared to students taught through a traditional communicative method. The results showed that students taught through the genre method outperformed students taught in the traditional way on reading comprehension and attitude and behaviour. In another study in Thailand, Waugh, Bowering, and Chayarathee (2005) studied grade 6 Thai students who were taught through the cooperative learning method and compared their performance with students taught through a traditional teaching method. The study showed that students taught through a cooperative method did better in reading comprehension, and had better attitudes and behaviour with regard to learning English than those taught with the traditional communicative method.

Research on English Reading Comprehension and English Writing as well as on Attitude and Behaviour have been conducted during this decade (2003-2013). Most of the research studies have drawn conclusions that bilingual students outperformed their counterparts. However, there was no study conducted to cover all the three variables, and in Aceh Province with Acehese context, as the present study did. Therefore, it was considered a need to conduct the present study in order to see how Acehese students

dealt with bilingual education as part of their challenge in education rehabilitation and revitalisation after the Tsunami 2004 disaster.

Gaps in Indonesian Research: Bilingualism *versus* Monolingualism

There appear to be four main gaps in the research literature in Indonesia (and especially Aceh) in regard to bilingualism and bilingualism education. One, while there is much anecdotal evidence in Aceh that bilingual schools are better than monolingual schools and that students taught bilingually perform better than students taught monolingually, there is no research evidence for this from Indonesia and Aceh. There is evidence from other countries supporting this, but not from Aceh itself. The judicial decision to ban bilingual teaching in Indonesia appears to have been made without any research evidence on the supposed benefits of bilingual teaching and learning. Secondly, there does not appear to have been any studies in Aceh or Indonesia involving experimental groups (with bilinguals) compared to control groups (with monolinguals). Thirdly, no relevant studies have been conducted in Indonesia (or Aceh) using modern measurement methods such as Rasch models to create linear scales of achievement and behaviour in regard to research relating to bilingually taught students. All the studies, whether relevant or not, have used True Score Theory (percentage scores) which are non-linear scores and items have not been created on a scale from easy to hard as is expected of linear measures. Fourthly, there does not appear to have been any research in Aceh which actually asks bilingually taught students what they think about bilingualism and bilingual teaching. No research appears to have been done to ask these students what benefits they think they are getting in learning English bilingually. The present research study aims to address these gaps (or deficiencies), at least partially.

The result of the present study would inform the education policy makers in Aceh about the current reality in Aceh that bilingual education was still needed, apart from the banning of bilingual education throughout Indonesia and the changes made to the current Curriculum 2013 in which the percentage of English lessons taught in primary and secondary schools in Indonesia schools was reduced. It would also suggest that the Acehnese government should reconsider the banning of bilingual education in Aceh Province and reestablish state international standard schools and state candidates

for international standard schools that have been banned due to the banning of bilingual education in Indonesia. As a province with a privilege to manage its own education policy, Aceh Province has the power to do this.

The next chapter explains the conceptual framework of bilingual and monolingual teaching, measurement, and the variables used in the present study.

CHAPTER THREE

CONCEPTUAL FRAMEWORK FOR BILINGUAL/MONOLINGUAL TEACHING, MEASUREMENT, AND THE VARIABLES

This chapter presents the conceptual framework behind bilingual and monolingual teaching in Aceh province, Indonesia, and explains why bilingual education is expected to produce superior achievements over monolingual education after a two month experiment. It also presents the three dependent variables of the study, that is English Reading Comprehension, English Writing and Attitude and Behaviour with regard to learning English, and explains the conceptual structure of each variable. Then there is an explanation of the problems of True Score Theory measurement and why it is not used in the present study. A better alternative – Rasch measurement – is explained, along with the output from one of the best Rasch computer programs, RUMM 2030 (Andrich, et al., 2010). RUMM 2030 produces some excellent graphics and tabular data to support the creation of linear unidimensional scales.

Conceptual Framework for Bilingual and Monolingual Teaching in Banda Aceh

Bilingual programmes and monolingual programmes in Aceh share similar characteristics. The characteristics are: similar entry knowledge, the same curriculum, the same amount of time for classroom teaching and learning, similar classroom English activities, similar outdoor English activities, similar time spent on homework, similar textbooks used, similar teachers' ability and similar number of other subjects to studied within one academic year.

The only difference between the two program types is the medium of instruction. Bilingual programs use a majority of English in English lessons (about 50% English and 50% Bahasa Indonesia), while monolingual programs use majority Bahasa Indonesia with a little amount of English in the English teaching-learning processes, or even, use Bahasa Indonesia combined with a local language (in this case, Acehnese)

with a little amount of English, especially for students in remote areas where the local language (Acehnese) is more convenient to use in teaching any subject, including English. This difference in language use, Bahasa Indonesia or English, may contribute to different outcomes for English learning for both types of students. In terms of students' ability in three variables of this research, that is, English Reading Comprehension, English Writing and Attitude and Behaviour with regard to learning English over a two month time difference, it is expected that bilingual students will produce better results than the monolingual students in regard to English Reading Comprehension, English Writing, and Attitude and Behaviour.

In regard to English Reading Comprehension, it would be reasonable to expect that bilingual students have a number of abilities in order to help them achieve higher scores in Reading Comprehension tests than their monolingual peers. First, bilingual students have more advanced phonological awareness than monolingual students (Bialystok & Herman, 1999). Not only is phonological awareness the most important of the metalinguistic skills that are basic to reading comprehension, but it also has repeatedly and independently been shown to underlie access to literacy and to ensure progress into fluent reading (Bialystok & Herman, 1999).

Second, being in a classroom with a second language instruction, bilingual students get benefits in interaction and communication with adult teachers. In this interaction and communication with teachers, bilingual students get models and response from their teacher to the students' meaning. It is strongly believed that when dealing with two languages, bilingual students tend to predict and hypothesise second language use and forms. They can expect, for example, word order, sentence structures, and a number of formal features of the utterances they use and hear (Cromdal, 1999). In line with this, the teacher would serve as model and could respond to language use and forms that the bilingual students have predicted and hypothesised. Being able to work out the predictions and hypotheses, the bilingual students also do self-correction. Self-correction, together with predictions, hypotheses and teacher's emphasis would enhance language use and forms, which bilingual students need and are capable of implementing.

In addition to that, the language interaction between teachers and bilingual students in a classroom would help students' auditory and visual perception and memory (Merisuo-Storm, 2007), and this is probably connected with learning a second or foreign language. In learning through a foreign language, bilingual students need to pay more attention to the second language. As a consequence, the students learn to listen with fuller attention than otherwise. They need to distinguish sounds, intonation, and stresses. This leads students to develop their auditory skills and visual perceptions. By also being able to know things in two names, bilingual students tend to memorise foreign words. Memorising foreign words helps develop the bilingual students' memories. These abilities are believed to help bilingual students to achieve better scores in Reading Comprehension than their monolingual counterparts.

In relation to English Writing, since bilingual students have literacy experience in both languages, it is expected that bilingual students would perform better in writing tests than monolingual students. Written language is a symbolic system in which letters represent sounds to indicate the phonological structure of a word, although this occurs less regularly in English than in Bahasa Indonesia. The letters are symbolic because they have no meaning and do not resemble the sound they represent (Bialystock, 1997). The bilingual classroom situations, in which students experience two different symbols for almost every subject in the environment, have enhanced the students' symbolic development. Students become familiar with many phrases and are capable of recognising written words. Being in a bilingual classroom with interaction and communication with a teacher and adopting the teacher's model, bilingual students can develop spelling skills that reduce spelling errors. They also tend to be able to control their grammar over anomalous sentences (Cromdal, 1999). These faculties are believed to help bilingual students to achieve better scores in writing tests than monolingual students.

In line with the learning of Reading Comprehension and Writing, it is important to note that bilingual students, especially those with low English proficiency might find it difficult to cope with English alone. In such condition, they might need to depend on the use of their first language to help them. When their proficiency increases, their dependence on their first language decreases.

Regarding the Attitude and Behaviour tests, it was expected that students in bilingual classes would have a significantly more positive attitude and behaviour with regard to studying English than their peers in the monolingual classes. It was expected that as the bilingual students showed their advanced ability in reading comprehension and writing tests, there would be a strong link to their confidence in reading and writing skills. The students who wrote and read well were the most successful in acquiring English. Consequently, the bilingual students had a significantly more positive attitude and behaviour with regard to reading and writing than the monolingual students. This is due to the expectation that bilingual students enjoy using English language more than monolingual students and that they are proud of their knowledge of and abilities in a foreign language. Being able to interact and communicate using a language that they like can awaken the students' interest in that language and, as a result, make them eager to further their command of English, and further improve their success in learning the subject matter as well. Positive attitudes towards language learning can raise motivation and help language learning.

To sum up, it would seem reasonable to expect that the bilingual situation in which a second language is used as the medium of instruction could affect not only students' ability of linguistic processing, but also their mental representation of the language. The explanation probably may be found in the students' ability to select and apply linguistic knowledge at will and thus, these abilities would keep them motivated. This would be expected to lead bilingual students to achieve better scores in reading comprehension, and writing tests, and to develop better attitudes and behaviour with regard to learning English than monolingual students.

It is expected that monolingual students would not benefit from their English monolingual program in regard to English Reading Comprehension, English Writing, and Attitude and Behaviour, as much as the bilingual students after two months of learning English reading comprehension and English writing. It is predicted that monolingual students would show less improvement in English, that is, English Reading Comprehension, English Writing and Attitude and Behaviour results, compared to those in the bilingual programme. By using Bahasa Indonesia or a local ethnic language (Acehnese), monolinguals do not have such a good exposure to English as those taught bilingually, nor do they spend much time on English tasks despite their lessons being of

the same duration. They do not employ metacognitive skills in English reading and writing as effectively, due to the expectation that they are not under pressure to do so. By having an inadequate English vocabulary, students find that English is not as important to them and not so challenging. This may lead to a poorer result in both English Reading Comprehension and English Writing scores. Feeling not so motivated towards learning English is expected to lead to a less positive feeling towards English. Therefore, it is predicted that monolinguals would show less improvement in their attitude and behaviour with regard to learning English after a two month time, compared to those taught bilingually.

Measurement

True Score Theory Measurement

True Score Theory Measurement, which is also called Classical Test Theory, is a theory that is based on the total correct score from a set of questions, which are not ordered by difficulty. This measurement is usually designed for particular subjects, such as English, Physics, History and so on (see Chapman, 2006; Keats, 1997). The model behind this measurement is that a 'true score' and a random error score from the particular set of questions and persons make the total correct score. This measurement is commonly considered to have at least six problems: First, non-linearity; second, multi-dimensional with 'noise'; third, item difficulties not ordered; fourth, person 'measures' and item difficulties are not ordered on the same scale; fifth, the 'measures' are item-content test dependent; and sixth, the 'measures' from different tests, even on the same topic, cannot be validly added or linked onto a single scale (see Michel, 1990; Smith, 1996; Waugh & Chapman, 2005; Wright, 1996).

Total scores are not considered linear due to the fact that equal differences between them do not represent equal amounts of understanding in whatever is being measured. In other words, the difference between 20% and 30%, and between 40% and 50%, and between 60% and 70%, for example, do not represent the same amount of understanding, or knowledge, as the case may be. While a total correct score usually represents a variety of knowledge, skills and understanding, this total score cannot accurately predict a person's response to questions in tests that consist of a set of questions, without considerable further analysis. It is also believed that total correct

scores contain a lot of ‘noise’ because there are no tests usually done in regard to unidimensionality, and also there are no tests usually done on consistency of student responses on an order of difficulty linear scale, that is, question difficulties are ordered on a continuum from easy to hard. For this study, the order of difficulty used is ‘easy’, ‘hard’ and ‘harder’.

Rasch Measurement

Contrary to the True Score Theory Measurement, in Rasch measurement, items are ordered from easy to hard on a continuum and their difficulties are calculated on a linear scale, which is also known as a log odds scale (see Andrich, 1988a; Andrich, 1989; Rasch, 1960). The person measures are calculated on the same linear scale. It is important to understand that when the data fit a Rasch measurement model, the differences between the person measures and the item difficulties can be calibrated together in such a way that they are freed from the distributional properties of the incidental parameter, because of the mathematics involved in the measurement model. This means that ‘scale-free’ measures and ‘sample-free’ item difficulties can be estimated with the creation of a mathematically objective linear scale with standard units. The standard units, called logits, are the log odds of successfully answering the items. Rasch measurement has been applied to many variables in education and educational psychology (see Waugh, 2003, 2005, 2010a; Waugh, 2010b).

The Simple Logistic Model of Rasch

The simplest Rasch measurement model for creating a linear scale was developed by the Dane, Georg Rasch (1901-1980) and originally published in 1960, with expanded editions later (Rasch, 1980, 2010). The Simple Logistic Model (SLM) of Rasch has two parameters: (1) representing a measure for each person on a variable and; (2) representing the difficulty for each item, although this is sometimes called the one-parameter model in the literature. The following are the characteristics of the Simple Logistic Model (SLM) of Rasch measurement (Rasch, 1960; Waugh, 2007).

- a. Items are designed to be conceptually ordered by difficulty along an increasing continuum from easy to harder for the variable being measured. For the purpose of explanation here, only three items are ordered from easy to medium to hard.

- b. In designing the items, one keeps in mind that person measures of the variable are conceptualised as being ordered along the continuum from low to high according to certain conditions. The conditions in this example are that persons with low measures will have a high probability of answering the easy items positively, and a low probability of answering the medium and hard items positively. Persons with medium measures will have a high probability of answering the easy and medium items positively, and a low probability of answering the hard items positively. Persons with high measures will have a high probability of answering the easy, medium and hard items positively. These conditions are tested through a Rasch analysis.
- c. Data were collected from persons on the items and scored dichotomously (0/1 or 1/2), as in, for example, but not limited to, wrong/right, no/yes, none/a lot, disagree/agree, some/often, bad/good, slow/fast.
- d. Each item is represented by a number, estimated from the data that represents its difficulty (called an **item parameter** in the mathematical representation of the Rasch Model) that does not vary for persons with different measures of the variable. Persons with different measures responding to the items have to agree on the difficulty of the items (such as easy, medium and hard, as used in this example). If the persons do not agree on an item difficulty, then this will be indicated by a poor fit to the measurement model, and then the item may be discarded as not belonging to a measure on this continuum.
- e. Each person is represented by a number, estimated from the data that represents his or her measure of the variable (called a **person parameter** in the mathematical representation of the Rasch Model) that does not vary for items of different difficulty along the continuum. If different items do not produce agreement on a person measure, then this will be indicated by a poor fit to the measurement model, and then one examines the person response pattern (and the items).
- f. Rasch measurement models use a probability function that allows for some variation in answering items such that, for example, a person with a high attitude measure may give a low response to an easy item, sometimes, or a person with a medium achievement measure might get a hard item right, sometimes. If the person response pattern shows too much disagreement with what is expected,

then it may be that the person has not answered the items properly or consistently, and that person's results may be discarded, or the item may be too hard or too easy, requiring it to be modified. In the mathematics of the model, the probability of answering correctly is related to the difference between the person measure and the item difficulty. In situations where there is a large positive difference between the person measure and item difficulty, then there is a strong probability of a correct response and, if there is a large negative difference, then there is a strong probability of an incorrect response. If the differences are not so large, the probabilities are changed appropriately (Waugh, 2007).

Equations for the Simple Logistic Model of Rasch

$$\begin{array}{l} \text{Probability of answering} \\ \text{positively (score 1)} \\ \text{for person n} \end{array} = \frac{e^{(B_n - \delta_i)}}{1 + e^{(B_n - \delta_i)}}$$

$$\begin{array}{l} \text{Probability of answering} \\ \text{negatively (score 0)} \\ \text{for person n} \end{array} = \frac{1}{1 + e^{(B_n - \delta_i)}}$$

Where:

e = natural logarithm base (e=2.7318)

B_n = parameter representing the measure (ability, attitude, performance) for person n

δ_i = parameter representing the difficulty for item i

Source: Rasch, 1960, 1980, 2010

These equations are solved from the data (entered in a text format) by taking logarithms and applying a conditional probability routine with a computer program such as RUMM (Rasch Unidimensional Measurement Models) (source: Andrich, 1988b; Rasch, 1960, 1980; Waugh, 2007).

The Extended Logistic Model of Rasch for Partial Credit Scoring

The Partial Credit Model of Rasch can be thought of as an extension of the Simple Logistic Model from two response categories to three or more response categories or outcomes (see Andrich, 1988b; Masters, 1988b; 1997). So the conditions, requirements and output of the Partial Credit Model are similar to the SLM, except that there are now more item parameters, and more item output and the equations are more complicated. The Partial Credit Model can be applied to any set of data scored, judged or answered in three or more ordered outcome categories where the level of outcome is conceptualised on a continuum from low to high.

Equations for the Partial Credit Model of Rasch

$$\begin{array}{l} \text{Probability of person } n \text{ scoring} \\ \text{in outcome category } x \text{ of item } i \\ \text{(for } x = 1, 2, 3, 4 \dots M_i) \end{array} = \frac{e^{\sum_{j=1}^x (B_n - \delta_{ij})}}{1 + \sum_{k=1}^{M_i} e^{\sum_{j=1}^k (B_n - \delta_{ij})}}$$

$$\begin{array}{l} \text{Probability of person } n \text{ scoring} \\ \text{in outcome category } x \text{ of item } i \\ \text{(for } x = 0) \end{array} = \frac{1}{1 + \sum_{k=1}^{M_i} e^{\sum_{j=1}^k (B_n - \delta_{ij})}}$$

Where:

e = natural logarithm base ($e=2.7318$)

$\sum (B_n - \delta_{ij})$ is the sum of $B_n - \delta_{ij}$

B_n = a parameter representing the measure (ability, attitude, skill or performance) for person n
 $\delta_{i1}, \delta_{i2}, \delta_{i3}, \dots, \delta_{iM_i}$ = are a set of parameters for item i which jointly locate the model probability curves for item i . There are M_i item parameters for an item with $M_i + 1$ outcome categories.

(Source: Masters, 1997, p.859)

Rasch Data Analysis with the RUMM (2030) Computer Program

There are eight data analysis tests (output) provided in the creation of a linear, unidimensional scale by using the Partial Credit Model of Rasch with the RUMM 2030 computer program. This output is similar for the Partial Credit Model of Rasch and the SLM of Rasch. The only exception is that for the Simple Logistic Model there are no ordered thresholds, only one threshold (*RUMM 2030 Manual*, 2009; Waugh, 2007).

a. Testing that the response category is answered consistently and logically

The RUMM program does this with two outputs: one, it calculates threshold values between the response categories for each item (where there are odds of 1:1 of answering in adjacent categories) and, two, it provides response category curves showing the graphical relationship between the linear measure and the probability of answering each response category.

b. Testing for dimensionality

An item-trait test-of-fit is calculated as a chi-square with a corresponding probability of fit. It tests the interaction between the responses to the items and the person measures along the variable and shows the collective agreement for all items across persons of different measures along the scale. If there is no significant interaction, one can infer that a single parameter for each person can be used to accurately predict each person's response to all the different items along the scale (described by a single parameter for each item) and it is in this sense that we have a unidimensional measure (Andrich & van-Schoubroeck, 1989, pp. 479-481).

c. Testing for good global Item-Person Fit Statistics

The item-person test-of-fit examines the response patterns for items across persons and the person-item test-of-fit examines the response patterns for persons across items using residuals. Residuals are the differences between the actual responses and the expected responses as estimated from the parameters of the measurement model. When these residuals are summed and standardized, they will approximate a distribution with

a mean near zero and standard deviation near one, when the data fit a Rasch measurement model.

d. Person Separation Index

The RUMM program calculates a Person Separation Index that is constructed from a ratio of the estimated true variance among person measures and the estimated observed variance among person measures by using the estimates of the person measures and their standard errors. This tests whether the standard errors are much smaller than the differences between the person measures (Andrich & van-Schoubroeck, 1989, p. 483).

e. Testing for good individual item and person residuals

Residuals are the differences between the observed values and the expected values estimated from the parameters of the Rasch measurement model. It is instructive to examine these outputs as they give an indication of whether persons are answering items in a consistent way and they give an indication of individual person and individual item fit to the measurement model.

f. Item Characteristic Curves

Item Characteristic Curves examine how well the items differentiate between persons with measures above and below the item location. An Item Characteristic Curve also shows a comparison between the observed and expected proportions correct for a number of class intervals of persons.

g. Person Measure/Item Difficulty Map

The RUMM program produces two types of person measure/item difficulty maps. These maps show how the person measures are distributed along the variable and how the item difficulties are distributed along the same variable (measured in logits). They show which items are easy, which ones are of medium difficulty and which ones are hard. They show how well the item difficulties are targeted at the person measures. That is, they show whether the items are too easy or too hard for the persons being

measured and whether new items need to be added, or whether there are too many items of similar difficulty (some of which are thus not needed).

h. Testing for construct validity

If the items are conceptually ordered by increasing difficulty (downwards) and the perspectives are ordered by increasing difficulty (to the right), then this represents the structure behind the variable. In Rasch measurement, all the item difficulties are calculated on the same linear scale and so the item difficulties can be compared with their conceptualised order. In this case, the item difficulties increase vertically downwards for each perspective by item and they increase horizontally to the right for each item by perspective. This provides strong support for the structure of the variable as it was postulated before the data were collected and analysed (Note: taken from Waugh, 2003; 2005; 2007; 2010a; and 2010b).

Variables

The present study measured three variables: English Reading Comprehension, English Writing, and Attitude and Behaviour with regard to English learning.

Measuring English Reading Comprehension

In this study English Reading Comprehension consisted of 11 multiple-choice questions and two essay questions. The multiple-choice items were arranged in order of predicted difficulty and are presented in four sub-parts: 'Needs for Energy' (easy); 'Sunlight and Producers' (hard); 'Consumers and Decomposers' (harder); and 'Humans in the Food Chains' (harder). Under each sub-part, three questions were also ordered by difficulty. Low question numbers under each sub-part were designed to be less difficult than high question numbers under the same sub-part. Thus, for sub-part 'Needs for Energy', as an example, question number 1 was considered easy; question number 2 was considered hard; and question number 3 was considered harder. This design applied to the other three subparts. The reason was due to the expectation that the test was designed in such a way that the low question numbers were easier for the students to answer; the higher the question number, the more difficult the question was for students to answer. The prediction that a question is considered easier than another was based on Bloom's taxonomy levels of thinking and thus whether the students were expected to

find it easy or hard to understand and/or consume less time in finding the correct answer. The opposite was that when it was hard for the students to locate the correct answers from the text, or if it took some time for them to find the correct answer from the text or it needed a higher level of understanding, then the items were predicted to be harder.

Take sub-part ‘Need for Energy’ as an example. Question number 1 (‘What will happen if there is no food?’) is considered the least difficult of all, because the question is based on common knowledge. The students were expected to understand it easily as long as they understood the question. For question number 2 (‘From where do plants get energy?’), the students need to read the text carefully and to understand it well, even though the answer was clearly stated on the text. Therefore, it was considered harder than question number 1. Question number 3 (‘A food chain is...’) was also stated in the text but students needed to have a higher level of thinking (as in Bloom’s taxonomy) to understand the concept in order to be able to answer the question correctly. Thus, it was considered the hardest of all. Other sub-parts followed the same pattern. To avoid repetition, other sub-parts are not discussed here (refer to Appendix A for the complete English Reading test and Appendix B for English Reading Comprehension’s ordered scoring scheme). Items (correct answer and distracters) for each question have been ordered by difficulty as well, in terms of three degree of correctness: ‘Most Correct’ (Score 3), ‘Partly Correct or Partly Incorrect’ (Score 2), and ‘Least Correct’ (Score 1) (see Table 3.1).

Table 3.1
Item Difficulty Scoring Scheme for Reading Comprehension Test

Reading Comprehension Item		Least correct	Partly correct or partly incorrect	Most correct
		Score 1	Score 2	Score 3
Item difficulty				
Need for Energy				
1	What will happen if there is no food?	a	b	c
2	From where do plants get energy?	c	a	b
3	A food chain is	a	c	b
Sunlight and Producers				
4	What do plants use to grow?	c	a	b
5	What do we call plants?	a	b	c
6	Plants make	b	a	c
Consumers and Decomposers				
7	What are primary consumers?	b	c	a

8	What is an example of a carnivore?	b	a	c
9	Decomposers are	c	a	b
Humans in the Food Chain				
10	What do humans eat?	b	a	c
11	What do we call humans?	b	c	a
12	Is the Food Chain important? Why? Or Why not? Write your answer below.			
	Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.			

Source: Designed by Khairiah Syahabuddin for the present study (see Appendix A for the complete Reading Comprehension test)

There were two Essay questions given to the students on the Reading Comprehension test sheets. One question, which is question number 12, was related to the Reading Comprehension text, and another one, which is without numbering, was related to their overall Reading Comprehension learning. Question number 12 was marked as part of the Reading Comprehension test, but, the final question was asked to allow students to provide some further comments on learning English Reading Comprehension. The question was not marked for scaling and was not analysed as part of the Rasch measure. One student's answer from the Reading Comprehension test number 12 is shown in Appendix N.

Measuring English Writing Achievement

The English Writing Test consisted of two compulsory topics, which were 'My *Idul Fitri* Holiday' and 'My Family' (see Appendix C for English Writing test). Students were asked to write some paragraphs about the two topics. They were encouraged to write as much as they could within the assigned time of 70 minutes for both topics. They were allowed to choose to write either topic first. Both writing topics were marked using the scoring rubric which was designed by the researcher. The scoring rubric for English Writing involved categories conceptualised from easy to hard and harder still (see Table 3.2). The scoring rubric has three main sections: Paragraph(s) Organisation, Text Conventions, and Text Quality. Each of these has three levels of predicted difficulty from 'easy', to 'hard' to 'harder'. The reason behind the scaling difficulty of the content is that it is expected that measurement implies an ordering of items by difficulty. It is expected that Paragraph(s) Organisation is easier than Text Conventions and Text Quality, because Text Convention involves aspects within paragraphs. This is also true for Text Quality. It is predicted that students find

Text Quality the hardest of all, because it encompasses the other two (Text Conventions and Text Quality).

Paragraph(s) Organisation contains three sub-parts: ‘Topic Sentence’, ‘Concluding Sentences’, and ‘Supporting Sentences, respectively. Text Conventions also have three levels of predicted difficulty from ‘easy’, to ‘hard’ and ‘harder’. They contain ‘Spelling’, ‘Punctuation/Capitalisation’ and ‘Grammar’ respectively. Text Quality also has three levels of predicted difficulty from ‘easy’, to ‘harder’ and ‘harder still’. It contains ‘Readability’, ‘Style’ (which includes Sentence fluency, such as varied length, good flow and rhythm, and varied structure), and ‘Text Organisation’, respectively. Each main section, which includes Paragraph(s) Organisation, Text Conventions, and Text Quality, has four categories of scoring: ‘Excellent’ for score 4, which is the highest score, ‘Good’ for score 3, ‘Adequate’ for score 2, and ‘Poor’ for score 1. The meaning of these is given in Table 3.2.

Level of difficulty for each of the three subparts follows the prediction of abilities needed by students to complete the tasks satisfactorily. For example, the Paragraph(s) organisation contains three subparts in order of difficulty: ‘Topic Sentence’ (easiest), ‘Concluding Sentences’ (hard), and ‘Supporting Sentences’ (harder). It is predicted that weaker-ability students find it easier to write a ‘Topic Sentence’, which might be regarded as simple knowledge, than to write a ‘Concluding Sentence’, and ‘Supporting Sentences’, which require a higher order of thinking (see the order of difficulty in Bloom’s taxonomy from knowledge to application to analysis to evaluation). Similarly, students are expected to find writing ‘Supporting Sentences, the hardest compared with the other two (‘Topic Sentence’ and ‘Concluding Sentences’). The reason is weaker-ability students are expected to be able to write a topic sentence, but not necessarily be able to write a concluding sentence, which requires a higher order of thinking, as described in Bloom’s taxonomy. Similarly, they might not produce some adequate and correct supporting sentences to support the main idea within the topic sentence. The Supporting Sentences were considered the hardest of all due to the expectation that the students needed more practice and knowledge to be able to write them than to write a topic sentence and a concluding sentence, and because they involve higher orders of thinking like analysis and synthesis.

Table 3.2
English Writing Scoring Rubric

		TEXT WRITING	EXCELLENT (Score 4)	GOOD (Score 3)	ADEQUATE (Score 2)	POOR (Score 1)
Paragraph(s) Organisation						
Easy	1	Main Topic	Main topic is clearly stated in a complete opening sentence with correct sentence structure.	Main topic is stated in an opening sentence but sentence structure is not correct AND/OR lacking detail.	Main topic is not clearly stated but sentence structure is correct.	Main topic is not clearly stated. Sentence structure is not correct.
Hard	2	Concluding Sentence	Concluding sentence restates topic with correct sentence structure.	Concluding sentence restates topic but sentence structure is not correct AND/OR lacking detail.	Concluding sentence is not clearly stated but sentence structure is correct.	Concluding sentence is not clearly stated. Sentence structure is not correct.
Harder	3	Supporting Sentences	Text has at least 3 detailed supporting sentences AND all sentences are on topic.	Text has at least 3 supporting sentences *Only 2 sentences are on topic. AND/OR *Sentences lack details	Text has at least 3 supporting sentences *Only 1 sentence is on topic. *Sentences lack details.	Text has supporting details BUT none are on topic.
Text Conventions						
Easy	4	Spelling	There are 0-2 spelling errors.	There are 3-4 spelling errors.	There are 5-6 spelling errors.	There are more than 6 spelling errors.
Hard	5	Punctuation/ Capitalisation	Text has 0-2 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 3-4 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 5-6 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has more than 6 errors in punctuation/ capitalisation, and noun-verb agreement.
Harder	6	Grammar	Text has 0-2 errors in noun-verb agreement.	Text has 3-4 errors in noun-verb agreement.	Text has 5-6 errors in noun-verb agreement.	Text has more than 6 errors in noun-verb agreement.
Text Quality						
Easy	7	Readability	Text is neat and readable with 0-2 marked out words or other corrections.	Text is neat and readable with 3-4 marked out words or other corrections.	Text is not neat and readable with 5-6 marked out words or other corrections.	Text is not neat and unreadable with numbers of marked out words or other corrections.
Hard	8	Style (Sentence fluency, e.g. varied length,	Text shows sentence fluency.	Text shows reasonable sentence fluency.	Text shows minimal sentence fluency.	Text lacking in sentence fluency.

		good flow rhythm, and varied structure)				
Harder	9	Text Enjoyability	Text is enjoyable to read	Text is quite enjoyable to read	Text is satisfactory to read	Text is not enjoyable to read

Source: Designed by Khairiah Syahabuddin for the present study (see Appendix C for the complete English writing test)

Questionnaire on Attitude and Behaviour with regard to Learning English

The Questionnaire on Attitude and Behaviour with regard to Learning English consisted of 18 statements and one essay question. The 18 statements are presented under seven sub-headings: Tasks for Listening, Tasks for Speaking, Tasks for Reading, Tasks for Writing, Student/student relationships, Student/teacher relationships and Common Views. Statements under each sub-heading have been ordered by difficulty, that is, easy, harder, and harder still. Take, for example, ‘Tasks for Listening’. It has three statements. Statement number 1 was considered easy; statement number 2 was considered hard; and statement number 3 was considered harder. Statements under the other sub-headings follow the same pattern. The reasoning behind this order relates to how hard it is expected for a student to hold each Attitude statement and how hard it is expected for a student to behave according to each Behaviour statement.

The arrangement of predicted items difficulties is in line with previous attitude/behaviour type Rasch measures performed by Waugh (Waugh, 2003; 2005; Waugh, 2010a; 2010b). Agreement indicated support for the structure of the variables. Each of the 18 statements was to be answered in two perspectives: ‘This is what I wish to happen’ (that is, Attitude) and ‘This is what does happen’ (that is, Behaviour), making an effective item total of 36. The Attitude statements under ‘This is what I wish to happen’ were considered easier than the Behaviour statements under ‘This is what does happen’ due to the expectation that to actually do ‘things’ at a high category level requires more effort than to merely think about what ought to be done ideally. Take for example statement number 1, ‘I pay attention to someone speaking English’. Under the Attitude perspective, one just thinks about how one ideally pays attention to someone speaking English; while under the Behaviour perspective, one has to make an effort to pay attention to someone speaking English and, because there are other distractions, one

has to focus and concentrate on the behaviour itself rather than just thinking about it. Under each perspective, the statements were to be answered in three ordered responses: ‘Most or all the time’ (score 3), ‘Some of the time’ (score 2), and ‘Never or rarely’ (Score 1). The single written-answer question at the end of the Attitude/Behaviour Questionnaire was designed to give an opportunity for students to express any other feelings or concerns regarding their attitude and behaviour with regard to learning English as a foreign language (see Appendix E for the full Attitude/Behaviour questionnaire). The essay question was not marked for scaling but was analysed qualitatively for common perspectives (details are explained in Chapter Eleven).

Table 3.3
Attitude and Behaviour Questionnaire

Questionnaire Item	This is what I wish to happen			This is what does happen		
	1	2	3	1	2	3
Item no.						
Tasks for Listening						
1-2 I pay attention to someone speaking English						
3-4 I ask others to speak slowly or repeat words in English						
5-6 I listen to English songs						
Tasks for Speaking						
7-8 I say new words several times in English						
9-10 I practise English with other students						
11-12 I start conversation in English with my friends						

Source: Designed by Khairiah Syahabuddin from several sources for this study. The full Attitude/Behaviour questionnaire is given in Appendix E.

The next chapter explains the Research Design for the present study.

CHAPTER FOUR

RESEARCH DESIGN

This chapter explains the design of the study. It covers the research strategy and design; mixed-method design; samples; pilot study; study and ethics approvals; control of extraneous variables in the quasi-experiments; test data collection, data entry, and data analysis of the quantitative and qualitative data.

Strategy and Design

This study involves a mixed-method design (Bergman, 2008; Clark & Creswell, 2008; Creswell & Clark, 2006) with two parts. The first part was a quasi-experiment with two intact groups involving a control group taught through monolingual instruction and an ‘experimental’ group taught through bilingual instruction, with pretest and posttest measures for English Writing, English Reading Comprehension, and Attitude/Behaviour with regard to Learning English. The data from both the pretests and posttests were analyzed using the computer program Rasch Unidimensional Measurement Models (RUMM2030) (Andrich, et al., 2010) in order to create linear measures. The linear measures were used with Mixed Design ANOVA, also known as Split-Plot ANOVA (with SPSS) with the pretest as the covariate, and Two-Way ANCOVA (with SPSS) (Pallant & Tennant, 2007).

The second part was the qualitative students’ written comments regarding their perceptions about learning English as a foreign language. The comments were taken from the same students who sat for the quantitative tests (N=394 Year 7 students taught English at bilingual schools and N=386 Year 7 students taught English at monolingual schools). The data were analysed by the analytic induction method to produce some general propositions or main issues (Miles & Huberman, 1994).

Mixed-Method Design

Mixed-method research is a design in which two different approaches are combined to collect both quantitative data and qualitative data, although the emphasis on each may not be the same (Bergman, 2009; Clark & Cresswell, 2008; McMurray et al., 2004). Bergman (2009) defined it as a design with a combination of at least one qualitative and at least one quantitative component in a single research project or program. More specifically, Morse and Niehaus (2009) defined it as a scientifically rigorous research project, driven by the inductive or deductive theoretical drive, and comprised of a qualitative or quantitative core component with qualitative-quantitative supplementary component(s). While the core component of the project is the complete method used to address the research question, the supplementary component is not. As it is a methodological strategy different from the method which is used to extend the investigation, the supplementary component is incomplete in itself or lacks some aspect of scientific rigour. It cannot stand alone and is regarded as complementary to the core component (Morse & Niehaus, 2009).

The mixed methods approach has three characteristic features: First, use of qualitative and quantitative approaches within a single research project; second, explicit focus on the link between approaches (triangulation); and third, emphasis on practical approaches to research problems (Denscombe, 2007). A mixed method design is believed to be stronger than one that uses only a single method, for example, one using only the quasi-experimental method (see for example, Bryman, 2006; Rocco et al., 2003; Collins et al., 2006; Greene et al., 1989). Descombe (2007) says that there are five reasons for this. The first is improved accuracy. Mixed methods design allows a researcher to use various methods for the same study, which improve the accuracy of the findings. Moreover, the design can assess the bias as well and it then becomes a valuable strategy for the development of research instruments. Greene et al. (1989) posit this use of a mixed methods approach can be used to “seek convergence, corroboration, correspondence of results from the different methods”. The second reason is that a mixed methods design gives a more complete picture of the answers to the research questions. The use of various designs with a mixture of data allows the researcher to have a description of matters under investigation. The third is compensating for strengths and weaknesses. Mixed methods design allows the researcher to be aware of

each method's strengths and weaknesses, which then give room to the researcher to recognise possible bias. By combining methods, it allows the researcher to compensate for the weakness. The fourth is developing the analysis. Mixed methods design allows the researcher to compare and contrast data better, which later can be used to better develop the analysis. The fifth is as an aid to sampling. The mixed method design allows the researcher to use information gathered as a basis to select respondents through a different and contrasting method (Denscombe, 2007).

Morse (1991) proposed two forms of methodology for mixed methods: (1) simultaneously, that is using both methods at the same time; and (2) sequentially, that is using the results of one method for planning the next method. She argues that using quantitative and qualitative methods to address the same research problem leads to the issue of weighing each method and their sequence in the study. The current study considers quantitative and qualitative methods simultaneously. Figure 4.1 illustrates the design of the study.

The present study took these comments into account and used a mixed-method approach involving a quasi-experimental design and the collection of some written comments from students at the posttest stage. The quasi-experimental design involved a control group (monolinguals) and an experimental group (bilinguals) with pretest and posttest measures. In order to gain some extra information students were asked to provide some written comments at the time of the posttest. These were analysed qualitatively and helped to provide a more complete answer to the research questions.

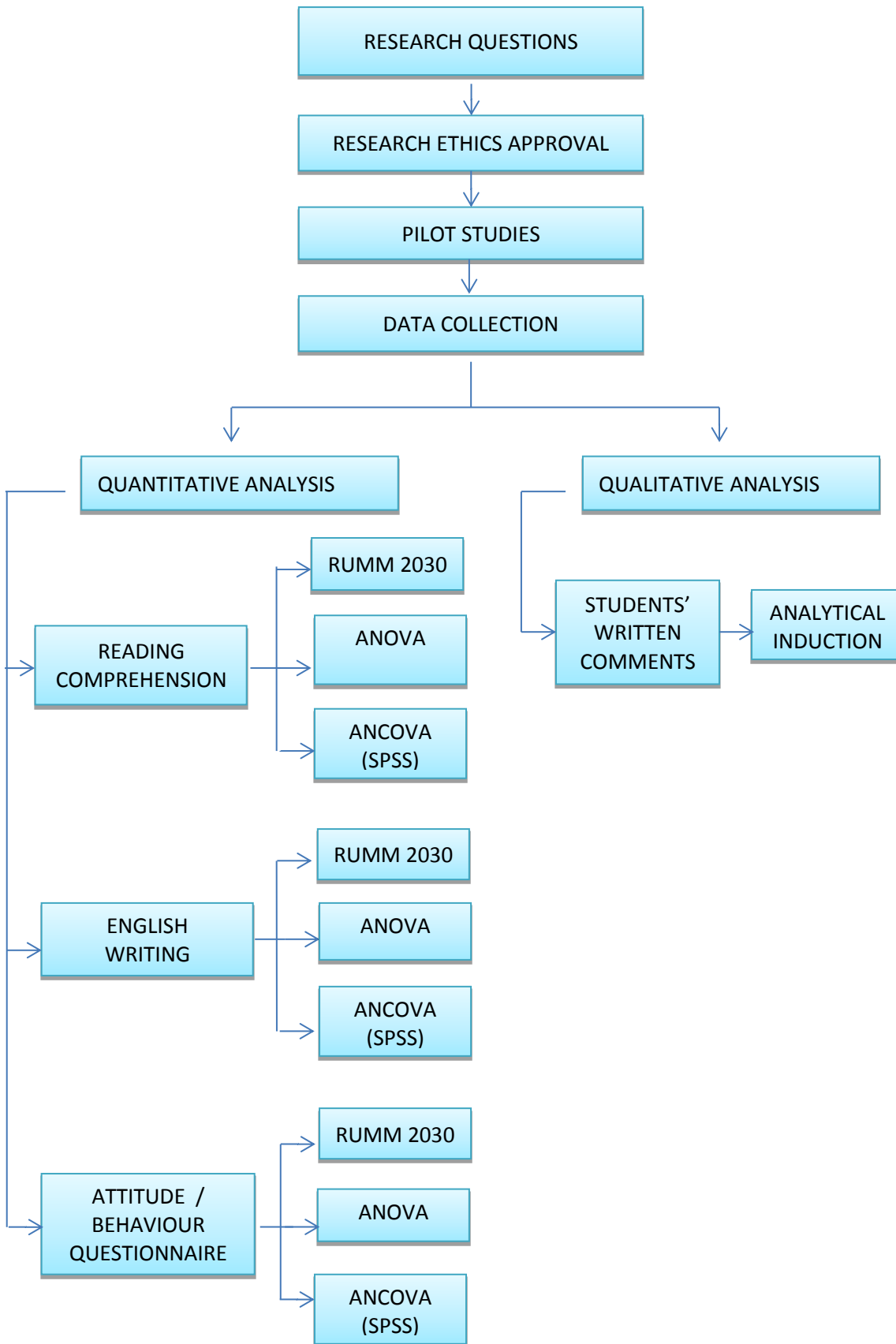


Figure 4.1 Design of the Study

Samples

Seven hundred and eighty male and female students who were enrolled in their first-year of middle schools (12/13 years old), consisting of 394 students from bilingual schools and 386 students from monolingual schools, were respondents for this study. The 780 students were taken from 13 State Middle Schools which are called 'SMPN'. 'SMPN' stands for '*Sekolah Menengah Pertama Negeri*', equivalent with Year 7 Primary School in the Australian context regarding the students' age. The schools are numbered under districts or suburbs: five schools having combined English and Bahasa Indonesia as medium of instruction, i.e. SMPN 1 (N=76), SMPN 2 (N=74), SMPN 6 (N=152), SMPN 13 (N=31), and SMPN 19 (N=61); and 8 schools using Bahasa Indonesia as a medium of instruction, i.e. SMPN 3 (N=76), SMPN 4 (N=48), SMPN 7 (N=50), SMPN 10 (N=52), SMPN 11 (N=19), SMPN 14 (N=33), SMPN 17 (N=63), and SMPN 18 (N=45).

The schools were taken from the same district or from neighbouring districts within one city, Banda Aceh. This was believed to be important to ensure that the students shared similar characteristics in terms of English teaching and learning experience; whether in bilingually-taught or monolingually-taught programs. Matching Bilingually-taught and monolingually-taught schools were chosen, as far as possible, by size and socio-economic status, but this also depended on voluntary agreement to participate. Schools surveyed for bilingual students were not also chosen to survey for monolingual students. It was intended that under no circumstances would students from either group interact with each other leading to collaboration between students.

The two groups shared the same characteristic regarding English learning, that is learning English as a foreign language. The two groups had the same English curriculum, similar syllabi, and they were taught with similar English teaching methodology. The students' after-school activities were similar. The major difference between them was the medium of instruction. Bilinguals were taught English as a subject using combined English and Bahasa Indonesia, while monolinguals were taught English using Bahasa Indonesia.

Almost all the students in both groups were of Acehnese (people of Aceh) origin and all spoke Bahasa Indonesia with their parents, siblings, school-mates and teachers.

The majority of them also spoke Acehnese as their heritage language with their parents, siblings, extended family, school-mates, and teachers. Only a few of them spoke English, read English books, watched TV programs, or were strongly involved in conversations in English with their classmates, outside the classrooms. English was rarely used in the monolingual classrooms but it was used in the bilingual classrooms. The only trivial difference between the two groups was the time for English lessons at school. The amount of English used for bilinguals varied, depending on the school, with a range of 38 hours to 57 hours per semester. The 38 hours per semester was made up of two hours for each meeting in a week for 19 meetings in a semester, while the 57 hours was made up of three hours for one or two meetings in a week for 19 meetings in a semester. The amount of English used for monolinguals was fixed, 38 hours per semester, consisting of two hours for one meeting for 19 meetings in a semester. Regarding this slight difference in the teaching-learning hours, the posttest for both groups was administered upon the completion of 32 hours of teaching-learning. This was considered important so as to ensure that both groups of students had been taught English as a subject for the same amount of time prior to the posttests being conducted.

Thus the samples were:

1. English Reading Comprehension and English Writing tests.
2. Questionnaire. All students who sat for the linguistic tests, that is, the English Reading Comprehension and English Writing tests, were requested to answer questionnaires on their attitudes and behaviour with regard to learning English.
3. Student Written Comments. All students who sat for the questionnaire test were requested to answer the last written item of the questionnaire which was about their experience in regard to learning English as a foreign language.
4. The 780 students were requested to sit for English Reading Comprehension and English Writing Experimental and Control Groups. Students who were taught English bilingually were assigned to the Experimental Groups, and students who were taught English monolingually were assigned to the Control Groups.

The study involved experimental and control groups with pretests and posttest using three measures: (1) English Reading Comprehension, (2) English Text Writing, and (3) an Attitude and Behaviour Questionnaire. Both control and experimental groups

were assessed under similar conditions and situations which involved the same English lesson content, similar lesson times and lesson duration, same homework, and same text books, but left the nature of the bilingual and monolingual teaching program unchanged. The only difference between the two groups was that the one group experienced bilingual teaching and learning processes with a slight extra amount of learning in class, while the second group experienced monolingual teaching and learning processes.

Teachers from bilingual schools and monolingual schools were counselled and monitored over the two months of the experiment to ensure that the controls on conditions and situations were working as they should. The points to consider were whether some changes in English lesson content, lesson times and lesson duration, homework, and text books occurred. Even though teachers in the bilingual and monolingual schools were different, they both shared similar awareness regarding the points of control that the study needed. The two-month experiment time was taken due to practicality issues and content issues. Practicality issues were teachers' willingness to participate in the research should the amount of experiment time exceed two months. Study issues were students' exposure to English language outside of the bilingual and monolingual classrooms. Since the focus of the study was to see the effect of bilingual and monolingual programs on student learning, controlling for extraneous English language variables outside the classroom was important. This was done by continual monitoring and counselling of teachers.

Table 4.1 shows the student numbers by control and experimental groups and by gender in control and experimental groups.

Table 4.1
Student Numbers by Control/Experimental Groups and by Gender in Control/Experimental Groups

Type of program	N	Gender	
		Female	Male
Bilingual	394	229	165
Monolingual	386	202	184

Instruments for English Tests

There are two researcher-designed English tests: (1) one for Writing English and (2) the second for English Reading Comprehension. No suitable tests could be found in the literature in line with the abilities of the Year 7 Banda Aceh students and so researcher-designed ones were used and piloted.

There are English Reading Comprehension Tests such as the ACER Neale Reading Comprehension Test (Neale, 1999) but this is just too advanced for the Banda Aceh students, so a researcher-designed one was prepared. The researcher-designed test involved students reading a passage about Food Chains taken from various sources in Aceh and then answering questions designed to test their comprehension of the text. There were twelve multiple choice items and two short written questions. There were three items on Sunlight and Producers; three items on Consumers and Decomposers; and two items on Humans in the Food Chains. One short written question was on the topic of the reading; while another one was a general question, which was on whether or not the students favoured Reading Comprehension. A scoring rubric was prepared so that the scored data was ready for Rasch analysis. Students were allowed to do the test within 40 minutes during school-time in school classrooms. The test was double-marked for consistency by the researcher and a pilot study was carried out prior to the main testing. Details of this test have been explained in Chapter Three (see Appendix B for English Reading Comprehension's ordered scoring scheme).

The English Writing Test contained two topics on which students wrote a number of paragraphs. The topics are 'My *Idul Fitri* Holiday' and 'My Family'. A scoring rubric was prepared in which three scoring aspects were ordered from easy to hard so that the scored data were ready for Rasch analysis. The scoring aspects are paragraph organization (topic sentence, concluding sentence, supporting sentences), text conventions (spelling, punctuation, grammar), and text quality (readability, style, and text enjoyability). The detail of this test has been explained in Chapter Three (see Appendix D for English Writing's ordered scoring scheme).

Questionnaire on Student Attitude/Behaviour

The researcher-designed questionnaire was based on a previously used Attitude and Behaviour with regard to English Learning Questionnaire published by Waugh, Bowering and Chayarathee (2005). This questionnaire was designed with Rasch measurement in mind and data from Year 6 primary students in Thailand produced a very good fit to a Rasch measurement model. Data collected with the questionnaire had a Person Separation Index of $PSI=0.92$ and a Cronbach Alpha of $\alpha=0.93$, showing good reliability. That is, a good linear, unidimensional scale was created with data collected with this instrument in Thailand.

Some changes were made to be applicable to Banda Aceh students. Waugh, Bowering and Chayarathee's Attitude and Behaviour (2005) had five tasks: Tasks for Group Work, Tasks for Meaning, Tasks for Reading Comprehension, tasks for Student/Student Relationship, and Student/Teacher Relationship. The present study had ten themes (or main issues): (1) Tasks on Learning English as a foreign language, (2) Student-Student Relationships, (3) Student-Teacher Relationships; (4) Personal Views on the Benefits of Learning English, (5) Common Views on the Benefits of Learning English, (6) Confidence and Achievement in Learning English, (7) Learning English through Media, (8) Family Support in Learning English, (9) Obstacles in Learning English, and (10) Other Views. The present study had three tasks with the same task name used in the previous study, namely: Tasks for reading comprehension, Student/Student Relationships, and Student/Teacher Relationship. However, the statements were partially amended. An example was Tasks for Reading Comprehension. The statements for the tasks of the previous study were: "I like to solve the problems/puzzles in reading assignments"; "I can put the story into the correct order"; "I can complete cloze exercises with the correct words"; and "I can find the correct answers to the reading questions". The statements for the present study were: "I guess the meaning of the English words in the text"; "I read words carefully in English"; and "I can read English at home on my own".

After some modifications were made, the revised questionnaire consisted of 21 items, each answered in two perspectives, making an effective item sample of 42. The two perspectives were: (1) 'This is what I wish to happen' (an attitude) and (2) 'This is

what does happen' (a behaviour). The 21 items were set out in order of difficulty and consisted of six parts: three items under Tasks for Listening, three items under Tasks for Speaking, three items under Tasks for Reading, three items under Tasks for Writing, three items under Student/Student Relationships, three items under Student/Teacher Relationship, and three items under Common Views. The questionnaire was translated into Bahasa Indonesia prior to the administering of the survey and a pilot study was carried out prior to the main testing. A 40 minute period was allocated to the questionnaire. Details of this test have been explained in Chapter Three (see Appendix E for the Revised Attitude/Behaviour Questionnaire, that is, the Attitude/Behaviour Questionnaire for the present study). The written comments were back translated into English upon being presented into Chapter Eleven.

Instruments for students' written comments

For the students' written comments, students were asked a final question on the Attitude and Behaviour Questionnaire "Is there anything else you would like to add about your experiences in learning English?" which had been translated into Bahasa Indonesian. Thus, the answers to this question were provided in Bahasa Indonesian by the students. There were 1846 comment, all were in Bahasa Indonesian), written by the students from both types of schools: with 1110 written comments from bilinguals, and 732 written comments from monolinguals. All the written comments were grouped into 13 themes or concepts and carefully tallied, as the following: Tasks for Listening, Tasks for Speaking, Tasks for Reading, Tasks for Writing, Student/student relationship, Student/teacher relationship, Personal Views on Benefit of Learning English as a Foreign Language, Common Views on Benefit of Learning English as a Foreign Language, Students' Confidence and Achievement, Learning English through Media, Family Support in Learning English as a Foreign Language, Obstacles of Learning English as a Foreign Language, and Other Views. Some comments were stated as students' quotations.

Administration of the Instruments

Procedure

First, the students from both schools where English was bilingually-taught and where English was monolingually-taught answered, as pretests, a 12 item multiple-

choice reading comprehension test and two topics on which to write, and a separate attitude and behaviour questionnaire. Second, after a two-month time which was equal to 16 teaching-learning meetings, the same students sat for the identical tests, as posttests.

One research assistant was present on the days of the pretest and posttest administration at each classroom to ensure the student-respondents understood the instructions. The classroom teachers were allowed to be present on the days of the tests but were not allowed to interfere in the test process in order to maintain the validity of the students' answers. The language of communication was the student-respondents' native language (Bahasa Indonesian). The test took about 40 minutes to answer. After two months, the same students sat for an identical test, as a posttest, under the same conditions.

The written comments were taken from the posttest Questionnaire, not from the pretest Questionnaire, because it was assumed that the students could give answers better after they had studied English longer, in this case, after a two-month period. Out of the 780 students (bilinguals, N=394; monolinguals, N=386) who completed the Attitude/Behaviour Pretest, 702 students (90%) provided an additional written comment. There were no problems encountered in gathering these data.

All the students' written comments, which were written in Bahasa Indonesia, were carefully read, translated into English, back-translated to ensure the accuracy of the translation, and classified into ten themes. Some student comments contained a number of themes. Each theme of a student's written comment was entered into a Microsoft Word table in the form of tallies which then were counted and checked again to ensure that they were accurately recorded.

Pilot Studies

A pilot study was carried out in October and November 2010 in Aceh, Indonesia, regarding the English Achievement Tests, that is the English Reading Comprehension, the English Writing Test, and the Attitude and Behaviour Questionnaire, on two types of students: bilinguals (N=31) and monolinguals (N= 28). The students who participated in the pilot study were excluded from participating in the main study that ran from January 2011 until April 2011. Some changes were made as a

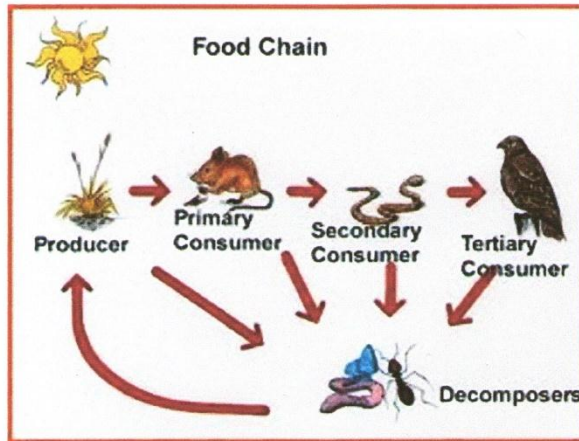
result of the pilot study for three variables: English Reading Comprehension, English Writing, and Attitude and Behaviour.

Pilot Study on English Reading Comprehension

A pilot study was carried out regarding this test. There were three changes resulting from the pilot study. First, the content and wording were changed, although the title remained unchanged. It appeared that the majority of the students found the text hard to understand. It is assumed that this was due to the text containing both complex and compound sentences and unfamiliar vocabulary. Therefore, it was changed to text containing a small number of complex and compound sentences and to more familiar vocabulary. For example the sentences like, '*Plants are called producers because they have the ability to convert the Sun's energy into chemicals that can be used to power life processes*' (Lines 4-5 of the original Reading Comprehension text) (see Appendix F for the original Reading Comprehension Text) were found to be difficult, since they contained a compound sentence (word 'and'), and complex sentence (word 'that'). The sentence of the text was then changed to '*All food chains begin with the sunlight. Plants use the sunlight to grow*'. The sentences were short, with repetitive words (*sunlight*) (see Appendix A for the revised English Reading Comprehension test, which was the English Comprehension test for the main data collection for English Reading Comprehension).

Second, a change was made to the pictures. Originally without pictures, the revised Reading Comprehension test contained colourful pictures. Pictures made the context clearer and more understandable for the students, especially for students with limited English in their initial year learning English as a foreign language. The following is an example of change in term of pictures. Figure 4.2 is the original Reading Comprehension Test Front Page, and Figure 4.1 is the revised Reading Comprehension Test Front Page.

FOOD CHAINS



Plants are important to our planet because they use photosynthesis to provide animal life with the energy they need to grow and to carry out life processes. When plants are eaten, this energy is transferred to another species. We call this a food chain.

There are three basic levels of a food chain.

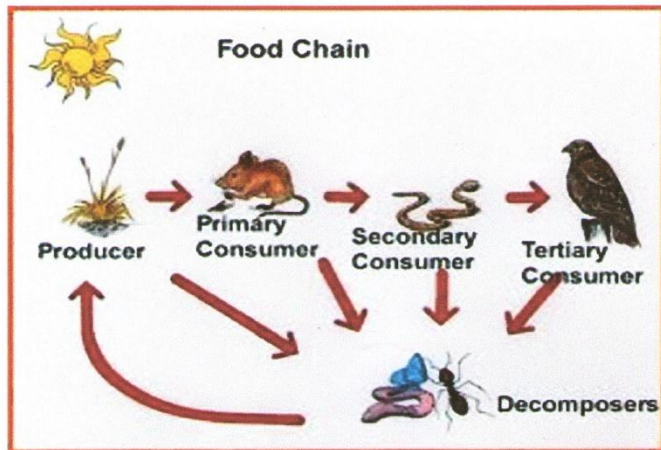
Level 1—Plants are called producers because they have the ability to convert the Sun's energy into chemicals that can be used to power life processes. Plants capture about 0.01 percent of the energy of the Sun's rays every day and are the basis of the whole food chain.

Level 2—Animals on the second level of a food chain are called primary consumers. We call them herbivores because they eat only plants for their energy. For example, deer eat shrub leaves, rabbits eat carrots and worms eat left litter. Primary consumers need to eat a lot of plants to meet their daily energy requirement. They eat approximately 10 per cent of the available plants and gain about 0.001 per cent of the Sun's original energy.

Level 3—Animals on the third level are called secondary consumers. They eat primary consumers for their energy, for example, lions eat deer, foxes eat rabbits and birds eat worms. Secondary consumers are good predators, but there are fewer of them because their energy sources are limited. Secondary consumers usually eat meat and we call them carnivores (if secondary consumers eat mainly insects, we call them insectivores). Secondary consumers gain about 0.0001 per cent of the Sun's original energy.

Further levels of the food chain are possible, but the longer a food chain is, the less energy left for the species at the top of the chain.

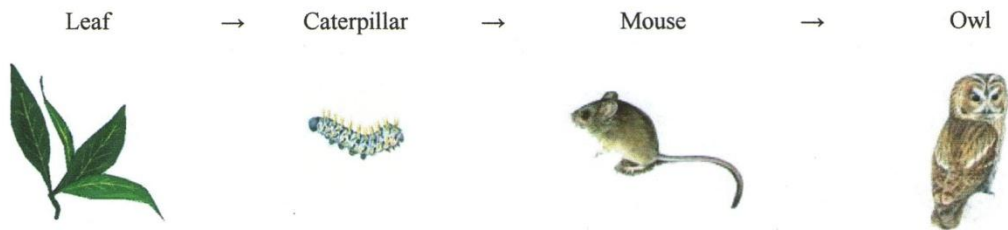
FOOD CHAINS



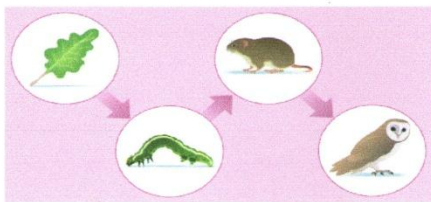
NEEDS FOR ENERGY

When you need energy in the morning before you come to school what do you do? Eat breakfast. Living things need food to give them energy. Plants get energy from sunlight. Caterpillars get their energy from plants. Mice get their energy from caterpillars. Owls get their energy from mice. Plants, caterpillars, mice and owls give energy to decomposers when die. The plants and animals in a food chain depend on each other to live. What would happen if all the plants died? The caterpillars would die. All the animals that depend on the caterpillars would die too. So every part of the chain is important.

Look at this picture representing a food chain using arrows,



Or this picture showing links or chains.



The caterpillar eats a leaf
 The mouse eats the caterpillar
 The owl eats the mouse

The third change was related to time. The test was originally allocated 40 minutes but this was changed to 45 minutes. The original time was scheduled so that the Reading Comprehension Test and the Attitude and Behaviour Questionnaire were administered at one meeting of 80 minutes, consisting of 40 minutes for the Reading Comprehension and another 40 minutes for the Attitude and Behaviour Questionnaire. The Pilot Study suggested that students needed a little more time for the Reading Comprehension Test, while they needed less time for the Attitude and Behaviour Questionnaire. To accommodate the students' need, the allotted time was changed for the main tests to become 45 minutes for the Reading Comprehension test and 35 minutes for the Attitude and Behaviour Questionnaire.

Pilot Study on English Writing

Changes were made to the topics. The original topics were 'My Holiday' and 'My Hobby'. Due to the observations carried out during the Pilot Study and due to the English teachers' comments in Aceh, the topics for the main data collection were changed to 'My *Idul Fitri* Holiday' and 'My Family'. There were two reasons behind this. First, it appeared that the students doing the Pilot Study did not have adequate vocabulary to answer the original topics sufficiently well. The students were in their initial semester of learning English and the Aceh teachers believed that the existing vocabulary learnt prior to the tests did not cover the original topics of the test.

Second, the pilot study showed that the majority of the students did not know what to write on the selected topics. This might mean that the context of the topics did not suit their daily context. In Acehnese, as well as in the Indonesian context, people do not have a specific time for holiday and hobbies, unlike in some other countries, like Australia, where people prepare and arrange their holidays during school holiday times. Therefore, due to the above reasons, the topics for the main data collection were changed to topics that were believed to be better understood and suited to the Acehnese context. The revised first topic 'My *Idul Fitri* Holiday' is believed as a correct choice, since *Idul Fitri* is the Acehnese biggest religious celebration, which is similar to Christmas in the Australian context. It is the most enjoyed time of the year. The second topic 'My Family' is well-known to the students, since they live with their family. Therefore, it was expected that the revised topics could give better opportunities to

students to express their beliefs and feelings, than the original ones (see Appendix C for the Revised English Writing test, which was the English Writing Test for the main data collection).

Pilot Study on Attitude and Behaviour with regard to Learning English

A pilot study was carried out regarding this test. The changes made were in three parts: (1) wordings of two statements; and (2) an additional category with three additional questions. Firstly, it appeared that some wordings contained ambiguity for the students and, therefore, they were changed to avoid misunderstanding. The wording that were changed was under the category ‘Student/teacher relationship’ with the wordings originally “I like my English teacher teaching English reading”, revised to “I like the way my teacher teaches English Reading Comprehension”; and with the wording, originally, “I like my English teacher”, revised to “I like the way my teacher teaches English Writing”.

Secondly, answered essay questions provided by some students doing the pilot study of the Attitude and Behaviour questionnaire suggested to the author that further questions were needed for a better understanding of attitude and behaviour with regard to learning English. Therefore, a new category was added, called ‘Common Views’. It had three statements: “I like English because we use it in the classroom”, “I like English because it helps me in higher study”, and “I like English because it helps me go abroad”. The full, original Attitude and Behaviour questionnaire is given in Appendix G, and the full revised Attitude and Behaviour questionnaire, which was the Attitude/Behaviour test for the main data collection, is given in Appendix E.

Ethical Considerations and Study Approval

This research followed the strict regulations of the Edith Cowan University’s (ECU) ethical guidelines and the Human Research Ethic Committee.

Ethics Approvals

An ECU Ethics Form was completed and sent to the ECU Ethics Committee for approval in relation to this proposal. The ethical guidelines of ECU ensured that the research was conducted in a fair and acceptable manner. The following section

describes some ethical issues under the ECU ethical guidelines.

Informed consent letter

Prior to the conduct of research, the Head of Provincial Education of Aceh was approached and requested to give permission for the researcher to conduct research at all potential Middle Schools in Banda Aceh. The headmasters, regular English teachers, and students were approached and asked if they would be willing to participate in this research. Four different consent letters were sent to these different recipients. All consent letters which were given to all recipients for this study were translated into Bahasa Indonesia. The first was to the Head of Provincial Education of Aceh, requesting that the research data collection be approved in all potential state Middle Schools in Banda Aceh (see Appendix H for Request Letter to Conduct Research Data Collection to the Head of Provincial Education of Aceh). The second was to the Headmasters of the Middle Schools in Banda Aceh. The consent letters containing the explanation of the research to all potential headmasters were personally delivered upon approval from the Head of Provincial Education of Aceh (see Appendix I for Consent Letter to the Headmasters). The third was consent letters that were sent to regular English teachers of the Middle Schools (see Appendix J for Consent Letter to regular English teachers). The fourth was consent letters to the potential students of Middle Schools in Banda Aceh (see Appendix K for Consent Letter to students doing Reading Comprehension, Appendix L for English Writing, and Appendix M for Attitude/Behaviour Questionnaire). These consent letters were attached to the paper test and were collected on the day of each test. Headmasters' permission was given on the condition that the research data collection was conducted at their schools. The participation of the regular teachers depended on their willingness to let their students do the test at the normal English teaching-learning session time. Students' participation depended on their agreement to sit for three different tests for both pretest and posttest.

Possible risks to participants

There were no anticipated risks to participants in the study and all participation was voluntary. The statement in the letter outlined the purpose of the experiment and ensured the regular English teachers and students of confidentiality and anonymity, with the right to refuse to participate, and to withdraw from the experiment at any time. After

they had read the information and were satisfied, they were asked to sign a form of consent. Then, students signed the form of consent based on the conditions mentioned above, indicating their willingness to participate.

Payment for participation

There was no payment involved in the research. The participants agreed to participate in the research without any compensation. Certificates of appreciation and participation were provided to the regular English teachers.

Study Approval

The proposal for the research was presented at a post-graduate seminar at Edith Cowan University with two reviewers. There were minor changes made and then the proposal was submitted formally through the Faculty of Education and Arts for approval.

After ethics approval and study approval were granted, the data collection was conducted twice in Aceh, Indonesia: for the Pilot Study, that took about two months and for the main data collection, that took about four months.

Control of Extraneous Variables in the Quasi-Experiments

Ideally, the experimental process should be free from problems that were potentially contributed by people and the process of the experiment itself. However, problems can occur occasionally during experiments. In order to minimize problems of extraneous variables in the quasi-experiments, the researcher did the following:

1. She visited all potential state middle schools in Banda Aceh and approached headmasters and English teachers and informed them about the study. They were informed that the study required students who were learning EFL, as a subject, which was taught either bilingually or monolingually. They were well informed that the nature and characteristics of English as a medium of instruction were crucial. Should English be used minimally, then the schools were labeled as monolingually-taught. On the contrary, schools that used mostly English or at least 50% English as a

medium in teaching English subjects were labelled as bilingually-taught. This point was made clear so as to avoid respondent recruitment and contamination of results.

2. The headmasters and English teachers clearly understood that the nature or characteristics of their students' learning were important. Each type of student to be recruited as respondent should share a similar learning nature and characteristics. Nature and characteristics here refers to how they learnt English in the classroom and at school. Did they practise English at school and at home? How well did they do that? Did they read English books/magazines/novels/comics which were written in English? Did they watch English movies? Did they listen to English music? Those questions were posed in order to ensure that both types of students shared the same nature and characteristics in learning English. The answers from the headmasters and the English teachers were confirmed as that both types of students shared a similarity of learning nature and characteristics. Should the nature and characteristics on the way students' learning English be obviously different, they could not be taken as respondents. This study required that each type of student shared the same learning nature and characteristics, so that the main difference was the use of English as the medium of instruction *per se*. Students who were taught English using at least 50% English as the medium of instruction would fall into 'bilinguals'; while students who were taught using Bahasa Indonesia or Acehnese with at the most 10% English as the medium of instruction, would fall into 'monolinguals'. Both the most engaged English students and the least engaged English students were part of the study. The most engaged English students were labelled 'bilinguals' and the least engaged English students were labelled 'monolinguals'.
3. English teachers agreed to monitor the students' learning nature and characteristics and to report to the researcher should any changes in the students' learning nature and characteristics occur during the two-month experimental period. The focus of the study was to investigate the effect of the use of combined English and Bahasa Indonesian (hence called the bilingual program) and the majority use of Bahasa Indonesian to English (hence called the monolingual program) to the students' English improvement (English reading comprehension and English writing). Controlling for the extraneous English language variables outside the classroom was important. Teachers were asked to report to the researcher by notes taken during the

experimental process any deviation from the controls. From day one until the last day of the experimental process, there were no changes reported.

4. The researcher kept monitoring and counselling the English teachers during the experimental process.

Test Data Collection, and Data Entry

Test Data Collection

Data for English Reading Comprehension, English Writing, and Attitude and Behaviour were each collected from 780 students from 13 state Middle Schools in Banda Aceh. Students in both bilingual and monolingual classes completed a comprehension test, a writing test, and an Attitude and Behaviour questionnaire for the Pretest. After two months they completed another identical Writing Test, Comprehension Test and Attitude and Behaviour questionnaire for the posttest. The tests were conducted during the English lesson teaching time. It meant that the time for the test was not arranged out of that particular class's English lesson. The test of Reading Comprehension and the Attitude and Behaviour questionnaire were taken at one meeting-period which was about 90 minutes: 45 minutes for the Reading Comprehension test, and another 35 minutes for the Attitude and Behaviour. The test of Writing was conducted on the following day. Example answers from student tests are given in Appendix N (for English Reading Comprehension), Appendix O (for English Writing), and Appendix P (for Attitude/Behaviour Questionnaire)

Data for the students' written comments (qualitative part of the study) were taken from part of the Attitude and Behaviour Questionnaire that was the final question which was required for respondents to answer. The written answer for this final question was done during the time for the Attitude and Behaviour test time. There was no chance for students to cheat, and the written comments were truly done by the students themselves, so the data maintained its accuracy. All data for the three tests were taken home for data entry and analysis straight after the test was done. There were no problems with collecting these data.

Data Entry

Test data collected were sorted. Only data from students who sat for both the pretest and posttest for the three variables were taken for the study. Students who sat only for one test (either pretest only or posttest only) were withdrawn from the study. Students who sat for both pretest and posttest for only one variable, for example, Reading Comprehension only, but did not sit for English Writing and Attitude and Behaviour, were withdrawn from the study. The data from the correct respondents were then carefully numbered for student codes, marked, and entered to Excel files. Checks and rechecks were conducted before they were satisfactorily ready for analysis to ensure that the data were entered accurately.

Qualitative data collected for the students' written comments were carefully read, and tallied into Microsoft Word for analysis, and checked for accuracy again.

Data Analysis

Analysis of Reading Comprehension and Essay Writing Test data

The data from both pretests and posttests were analyzed using the computer program Rasch Unidimensional Measurement Model (RUMM2030) (Andrich, et al., 2010) to create linear scales. This computer program provides various statistics that fit the measurement model and various graphs relating to the linear scale (item difficulties with standard errors, person measures with standard errors, item response category curves, item characteristic curves for discrimination and targeting graphs by context variables like gender). The RUMM computer program was used to create linear scales for use in identifying students' linguistic achievements for both bilingually-taught school students and monolingually-taught school students. The measures obtained from the RUMM computer program were then used in the Mixed Design ANOVA, also known as Split-Plot ANOVA (with SPSS); and the Two-Way ANCOVA (with SPSS) (Pallant & Tennant, 2007) in order to compare pretest and posttest linear measures by control and experiment groups. The posttest measures were taken as the anchor (i.e. the process of using anchor values, which is a preset logit value assigned to a particular test/measure to be used as a reference value for determining the measurements or calibrations of other tests/measures) so that both pretest and posttest measures were on the same posttest scale and could be validly compared.

The items of the Reading Comprehension Test were conceptualized in order of difficulty and the three responses to each item were also conceptualized in order of difficulty (see Appendix A for Reading Comprehension Test, and Appendix B for English Reading Comprehension's ordered scoring scheme). Since the Rasch analysis calculated the actual item difficulties on the same linear scale, it was possible to compare the conceptualized item difficulties and their measured item difficulties. This is a construct validity test of the structure of the variable.

An ordered scoring rubric was conceptualized for the Writing Test (see Appendix C for English Writing Test and Appendix D for English Writing's ordered Scoring rubric). It is based on the structure of a good text: Text Organization, Text Conventions, and Text Quality. The ordered scoring has criteria for each of the three parts: for excellent (score 4), good (score 3), adequate (score 2), and poor (score 1). Text Organization (topic sentence is easy; concluding sentence is hard; and supporting sentences are harder); Text Conventions (punctuation/capitalisation is easy; spelling is hard; and grammar is harder); and Text Quality (readability is easy; style is hard; and text organisation is harder).

Details of the Rasch data analysis for Reading Comprehension are given in Chapter Five; and the details of the data analysis for the experimental measurement of Reading Comprehension are given in Chapter Six.

Analysis of Attitude and Behaviour Questionnaire Data

The questionnaire data for students were used to identify students' attitude and behaviour with regard to the English language program at the schools with special reference to bilingual and monolingual schools. These data were analysed by using the computer program Rasch Unidimensional Measurement Model (RUMM2030) (Andrich, et al., 2010) to create a linear, unidimensional measure. The measures were used in the Mixed Design ANOVA and the Two-Way ANCOVA (with SPSS) (Pallant, 2007), involving pretest/posttest, and control/experimental group measures with SPSS.

The items for the Attitude and Behaviour Questionnaire have been conceptually ordered by difficulty under six sub-headings: (1) Tasks for Listening, (2) Tasks for Speaking, (3) Tasks for Reading, (4) Tasks for Writing, (5) Student/Student

Relationships, and (6) Student/Teacher Relationships. Each item was answered in two perspectives by the students: (1) 'This is what I wish to happen' (an attitude) and (2) 'This is what does happen' (a behaviour), with a set of ordered response categories: scored 1 for never or rarely, scored 2 for some of the time, and scored 3 for most or all of the time. This provides a conceptual structure of item difficulties that can be compared with the actual Rasch-created item difficulties on the same linear scale to test the structure of the variable (its construct validity), like a science experiment. Details from the Rasch Analysis of Attitude and Behaviour measurement are given in Chapter Nine; and the details from the data analysis of the experimental measurement of Attitude and Behaviour are given in Chapter Ten.

Analysis of Students' Written Comments

The students' written comments were carefully read, classified into themes, tallied, and analysed using the analytic induction method (Miles and Huberman, 1994). Ten themes (abstractions or propositions) related to attitudes and behaviour with regard to the English language program at the schools were identified from the student comments. Details from the data analysis of Students' Written Comments are given in Chapter Eleven.

The next chapter presents data analysis and findings from the Rasch measurement of Reading Comprehension.

CHAPTER FIVE

DATA ANALYSIS (PART 1)

RASCH MEASUREMENT OF READING COMPREHENSION

This chapter shows the findings of the Rasch analysis of the Reading Comprehension posttest data.

The Reading Comprehension Test (N= 12 items) was designed to measure the students' English Reading Comprehension at the posttest stage of the experiment. The data from the posttest were analysed with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010). The Reading Comprehension Posttest was scored in two categories of 0 (for wrong) and 1 (for correct) for 10 items, namely: items 1, 2, 3, 4, 5, 7, 8, 9, 10, and 11. These items were analysed using the Simple Logistic Model of Rasch (Rasch, 2010). Items 6 and 12 were analysed by using the Extended Logistic Model of Rasch because they had three scoring categories, i.e. 0, 1, and 2 (see Andrich, 1988a).

The English Reading Comprehension test was not part of the school assessment and was conducted for the purpose of this study. The test items were designed conceptually to fit on a continuum from easy to hard, and to be of the right difficulty range for the Indonesian students and to have all items gender and type of program (bilingual and monolingual) neutral. The test items were designed so as not to promote certain gender or type of program, for example, the bilingual program over the monolingual program, or vice versa. The idea behind the fit-on-continuum design is that students' ability could be predicted. It assumes that students with low ability would correctly answer the easy questions but would find it hard to answer correctly the moderately difficult questions and the difficult questions. The students with moderate ability would correctly answer both the easy questions and moderately difficult questions, but not the difficult questions; while students with high ability would

correctly answer the easy questions, the moderately difficult questions and the difficult questions.

The idea was to choose a text passage with questions of the right difficulty range and, at the same time, choose a neutral reading passage that was not beyond the students' ability (for both bilinguals and monolinguals) in terms of understanding the passage, but not favouring either female or male students. The chosen text passage was about The Food Chain, which was familiar to the students from their previous classroom work. This kind of passage had been taught in the students' previous lessons from year 1 to year 6 in Bahasa Indonesian and was considered important because the students learnt English for 4-5 months for one semester, with about 30-40 teaching hours within those 4-5 months. The majority of the students had very little, or no English at all, prior to their first year study in middle school. In Indonesia, English is taught as part of the curriculum in the first year of middle school, with an average of 38 teaching hours (1 teaching hour = 45 minutes).

Rasch measurement shows what should be expected in response to items if measurement at the metric, or linear level, is to be achieved (Pallant & Tennant, 2007). There are a number of statistical and graphical outputs to be considered using this analysis (see RUMM Manual, Waugh, 2003, 2005, 2010a, 2010b). These include Standardized Fit Residuals; Item-Trait Interaction (dimensionality); Person Separation Index, Individual Item Fit, Threshold Values, Response Category Curves, Item Characteristic Curves, Targeting, Differential Item Functioning (DIF) by Gender, Mean Measures by Gender, Differential Item Functioning (DIF) by Type of English programs, and Item Threshold Values.

The overall fit to the Rasch Model was not ideal. This meant that there was not good agreement about the item difficulties along the scale. Nevertheless, some other measurement aspects were satisfactory.

Output Analysis

Standardised Fit Residuals

Standardised Fit Residuals show the response patterns for items across students, and the response patterns for students across items. This interaction establishes the

overall fit statistics that confirms whether the item estimations contribute meaningfully to the measurement of one construct and sums over all items for each person and over all persons for each item. For an ideal well-targeted measure, which is not too easy or too difficult, the fit residual location mean is about zero and its standard deviation is about one (see Table 5.1) In this case, the fit is not ideal, but may be considered satisfactory when all the other RUMM output is considered.

Table 5.1
Overall Fit Statistics Reading Comprehension Measure (N=780, I=12)

	ITEM-PERSON INTERACTION			
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-1.59	1.54	-0.48
Standard Deviation	1.01	2.82	1.12	1.04

Notes:

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model.

Principal component analysis of residuals

The Principal Components analysis of the residuals showed that the item residuals loaded on a number of factors. However, deleting the not-so-well-fitting items did not improve the overall fit to the Rasch Model. The eigenvalue of the first component was 1.50 and that of the second component was 1.40, both of which are close to the chance level, supporting the finding of a unidimensional measure (see <http://www.rasch.org/rmt/rmt191h.htm>)

Item-trait interaction (dimensionality)

Item-trait interaction examines the consistency of agreement between students and the item difficulties (see Table 5.2). It checks that all the students agree that a particular item is easy, or of medium difficulty, or hard. In this case, the item-trait chi-square showed less than ideal agreement amongst the students on the difficulties of the items along the scale. It would seem that, in many cases, half of the students with medium measures said the item was easy and half said it was of medium difficulty. A similar situation occurred for some of the students with low measures.

Table 5.2
Item-Trait Interaction for Reading Comprehension Scale

Total Item Chi-Square	281.83
Separation Index	0.55
Total Degree of Freedom	72.00
Total Chi-Square Probability	0.00

Notes:

1. The item-trait interaction test is a chi-square. The results indicate that there was less than ideal collective agreement amongst the students about the item difficulties along the scale.
2. All numbers are given to two decimal points because the errors are only up to two decimal points.

The item-trait interaction chi-square was $\chi^2 = 200.4$, $df = 60$ with $p=0.00$ and, while five out of the 12 items fitted the model, they did not fit ideally well. It may be that some guessing of item answers by students at different measures along the scale was responsible for this non-agreement of the item difficulties, perhaps because of their low English Reading Comprehension standard and because of their desire to perform well. It seemed that there could be more than one scale present (or a number of factors or dimensions), but other data did not support this.

Person Separation Index

The Person Separation Index is an estimate of the true score variance among the students and the estimated observed score variance using the estimates of their ability measures and the standard error of these measures (Andrich & van Schoubroeck, 1989). For a good measure, the index is desired to be 0.75 or greater. The Person Separation Index, constructed as the ratio of the estimated true variance among the persons and the estimated observed variance among the persons, using the estimates of their locations and the standard errors of those locations, was $=0.55$, and is low. The Cronbach Alpha (Cronbach, 1951), based on the raw scores, was also low at 0.68. This indicates that the measures were only moderately well separated in comparison to the errors. It appears that there was insufficient variation in student measures along the scale in spite of the large sample size and, in retrospect; it seems that the 12 items were too similar and perhaps too similarly focused on the same content.

Individual Item Fit

Table 5.3
Locations, Standard Errors, residuals and Chi-Squares for Reading Comprehension Items

Item	Location	SE	Residual	ChiSq	DF	Probability
1	-0.545	0.122	-2.028	9.896	5	0.08
2	-0.118	0.077	-4.010	22.588	5	0.00
3	-0.932	0.140	-2.431	9.536	5	0.09
4	-1.098	0.149	-1.106	5.106	5	0.40
5	-0.296	0.113	-2.115	7.232	5	0.20
6	1.330	0.061	1.379	43.693	5	0.00
7	1.244	0.083	2.581	14.083	5	0.02
8	-0.848	0.136	-3.043	14.309	5	0.01
9	-0.505	0.121	-2.512	14.902	5	0.01
10	-0.341	0.115	-2.050	11.346	5	0.04
11	-0.749	0.131	-2.945	16.363	5	0.01
12	2.860	0.088	-1.917	45.087	5	0.00

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.
3. df means degrees of freedom. Probability is based on the chi-square fit to the measurement model.

The RUMM 2030 program calculates individual item fit to the measurement model. The fit statistics for all items for Reading Comprehension measure are presented in Table 5.3. While five items fitted the model, the other seven do not discriminate very well as expected from the model.

Raw residuals are the differences between the observed and expected responses. Standardized fit residuals are the differences adjusted to their standard deviations and they should be within the range $-2 < x < +2$. Table 5.3 indicates that some residuals are not ideal and that some items have not produced data that have a good fit to the measurement model.

Threshold Values

Table 5.4

Item Thresholds-uncentralised (Item=12, Number=780) for Reading Comprehension Measure

Item	Item Location	Mean Threshold	Thresholds	
			1	2
1	-.06	-.06	-.064	
2	-.45	-.45	-.454	
3	-.99	-.99	-.995	
4	-.38	-.38	-.379	
5	-.32	-.32	-.323	
6	+.98	.98	+.137	+1.833
7	1.34	1.34	1.337	
8	-.88	-.88	-.880	
9	-.49	-.49	-.488	
10	-.57	-.57	-.567	
11	-.52	-.52	-.519	
12	2.35	2.35	2.277	+2.418

Thresholds are points between nearby categories where the odds are 1:1 of answering in either category. The thresholds for items 6 and 12, which have three response categories (items scored 1, 2 or 3) are ordered, meaning that the response categories were used consistently and logically (see Table 5.4).

Scoring Category Curve

Scoring Category Curves represent the relationship between the probabilities of scoring in each category and person measures along the scale. Some items have two levels of scoring: zero for an ‘incorrect’ answer and one for a ‘correct’ answer for each item, and two items have three levels of scoring: zero is for an ‘incorrect’, one is for ‘partly correct’ and two is for a ‘correct’ answer.

Figure 5.1 is the Scoring Category Curve for item 1 of Reading Comprehension, which was scored dichotomously. The figure of its Scoring Category Curve shows that the scoring was done consistently and logically. When students have a low measure on item 1, then they have a high probability of answering item 1 incorrectly, thus receiving a zero score. When they have a high measure on item 1, they have a high probability in answering item 1 correctly, thus receiving score 1. The scoring for another ten items was checked and they were satisfactory too.

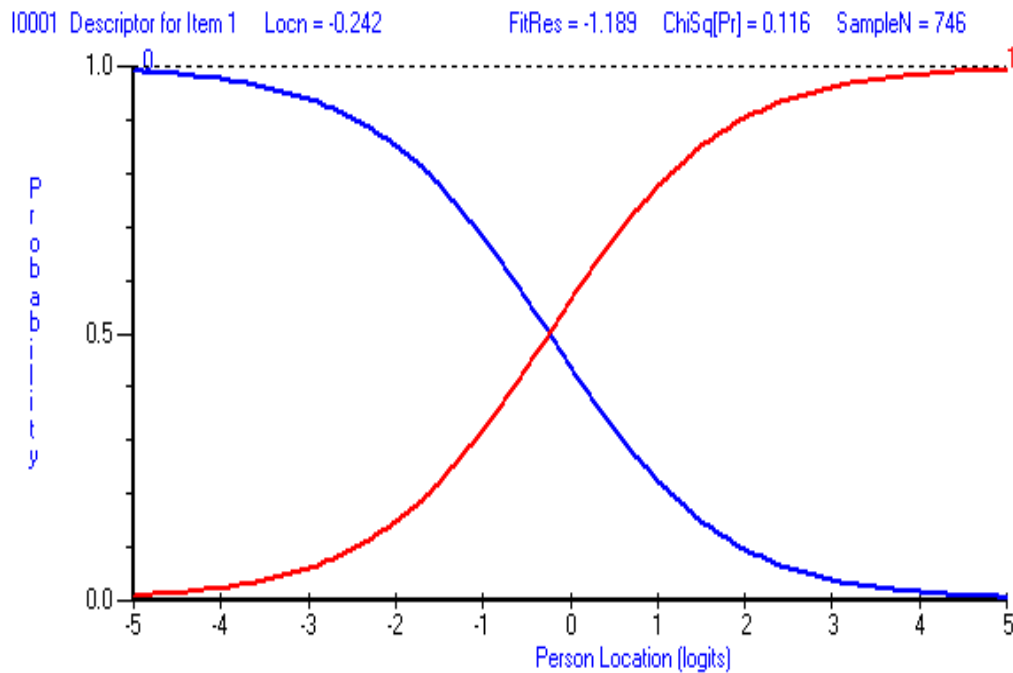


Figure 5.1 Item Category Curve for Item 1

Notes:

1. Blue colour curve shows the probability of scoring 0.
2. Red colour curve shows the probability of scoring 1.

Figure 5.2 is the Scoring Category Curve for item 6, scored 0, 1, 2. The figure of its Scoring Category Curve shows that the scoring was done consistently and logically. The Category 0 curve shows that if a Reading Comprehension Measure is located at -0.5 logits, the probability of answering the item correctly is around zero and, if the Reading Comprehension Measure is located at +5.0 logits, the probability of answering the item correctly is around 1. If a Reading Comprehension Measure is located at -5.0 logits, the probability of answering the item incorrectly is around 1. If the Reading Comprehension Measure is located at around +1 logits, the probability of answering the item partly correctly (and scoring 1) is around 0.5. If the Reading Comprehension measure is around +6.0 logits, then the probability of answering correctly is close to 1. The Category Scoring Curves for all the other 11 items were checked and they were satisfactory too.

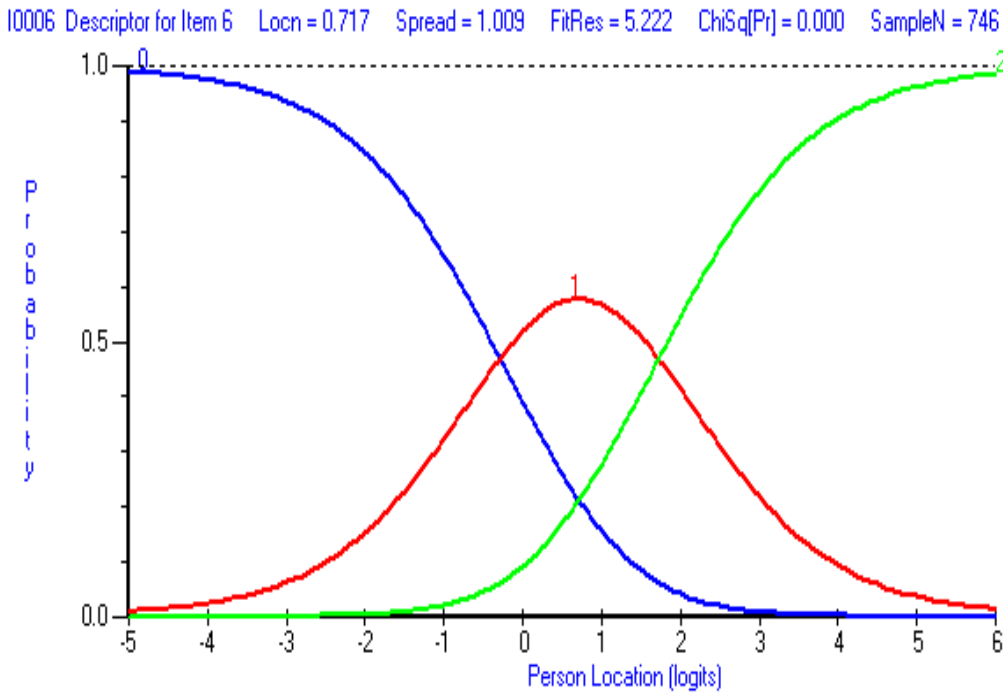


Figure 5.2 Item Category Scoring Curve for Item 6

Notes:

1. Blue colour curve shows the probability of scoring 0.
2. Red colour curve shows the probability of scoring 1.
3. Green colour curve shows the probability of scoring 3.

Item Characteristic Curves

Item Characteristic Curves show the relationship between the expected response score and the student measure, so that the researcher can examine how well the item differentiates between persons with measures above and below the item location. The observed means, shown as dots, in the seven class intervals are close to the ogive curve for item 1 (Figure 5.3) but not for item 6 (Figure 5.4) which shows that the item is not discriminating very well.

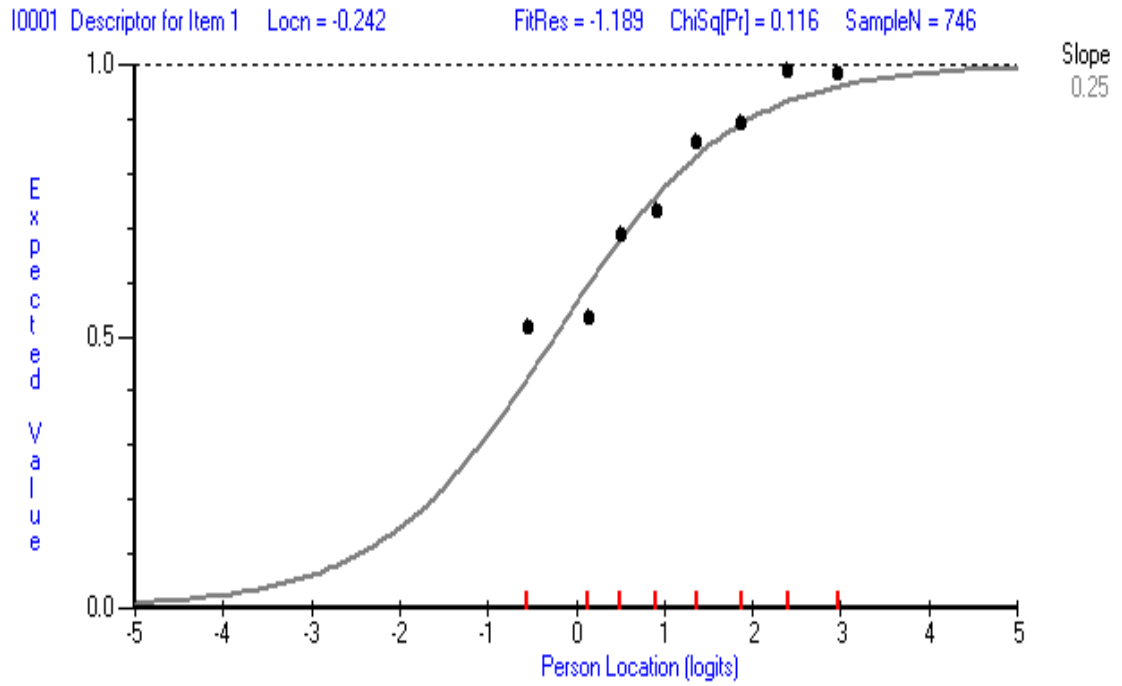


Figure 5.3 Item Characteristic Curve for Item 1

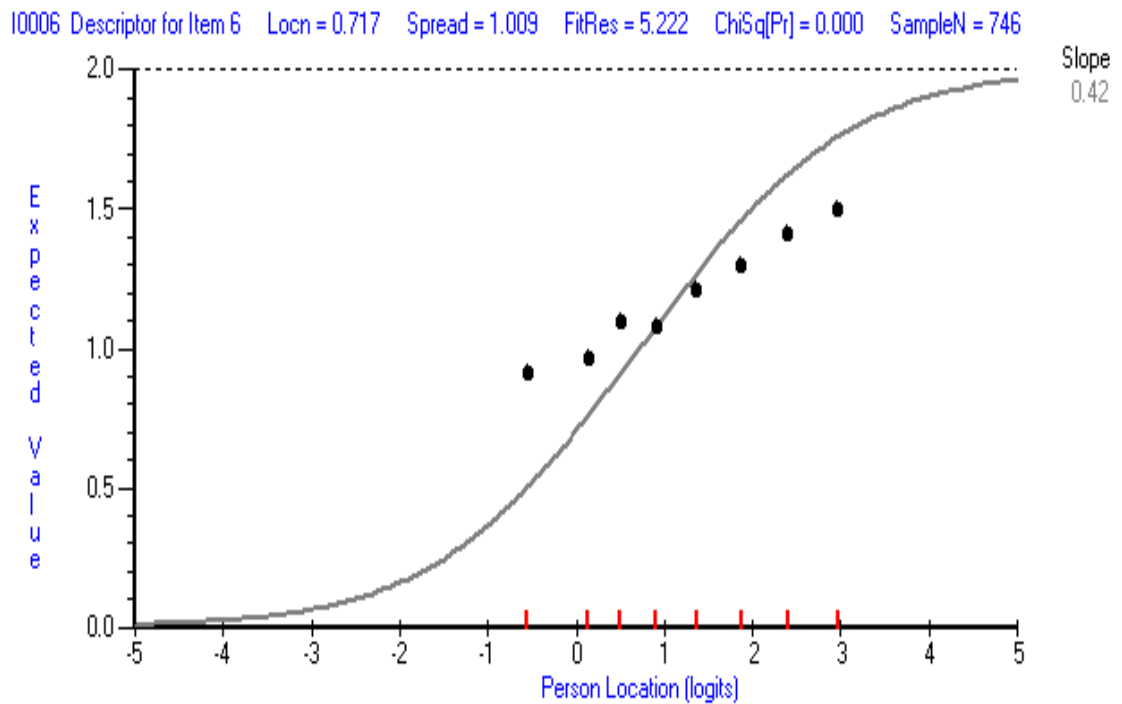


Figure 5.4 Item Characteristic Curve for Item 6

The Item Characteristic Curves were satisfactory in most cases without being ideal and targeting was reasonable – the student measures ranged from -0.8 to +3.9 logits and the item thresholds ranged from -1.2 to +2.6 logits.

Person-Item Threshold Distribution (Targeting)

Targeting shows how well the item difficulties are distributed along the scale against the student measures. A scale with a good measure should have a range of item difficulties that correspond to students with different student measures (Maley & Bond, 2011). Figure 5.5 shows student measures of Reading Comprehension from low to high on the top of the scale and item thresholds ranging from easy to hard on the bottom of the scale. The distribution graph showed that there were insufficient hard items to cater for the high measuring students and insufficient medium difficulty items to cater for the students with medium level measures. Before the scale is used in the future, more items should be added.

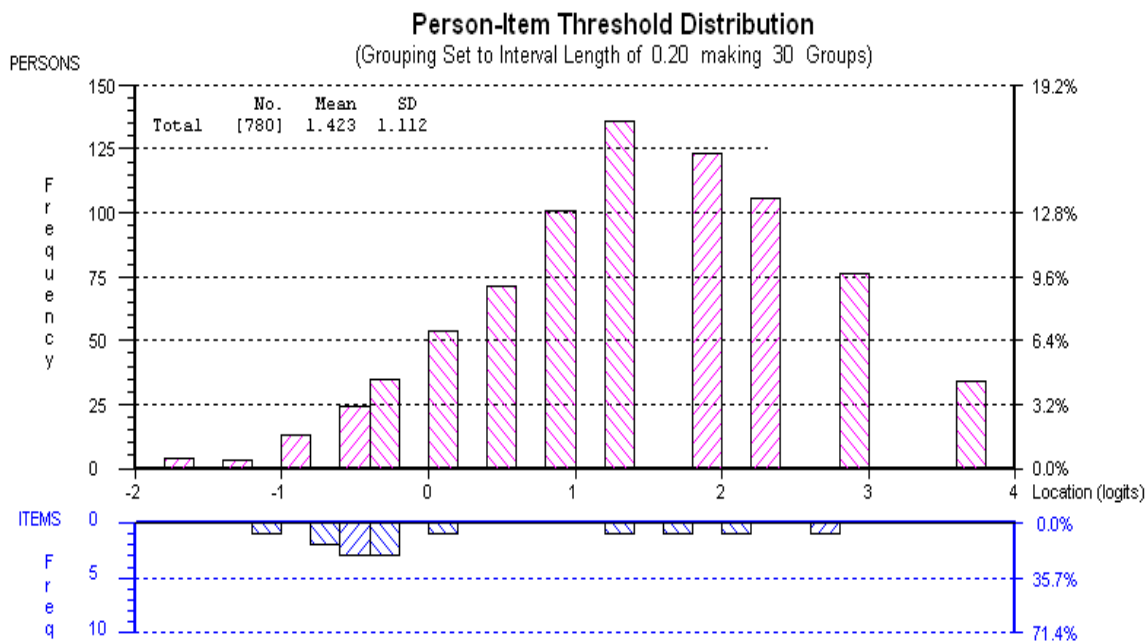


Figure 5.5 Person-Item Threshold Distribution for Reading Comprehension

The distribution of students and item difficulties are also illustrated in Figure 5.6, in which students' reading comprehension and item 'difficulties' were calibrated on the same scale.

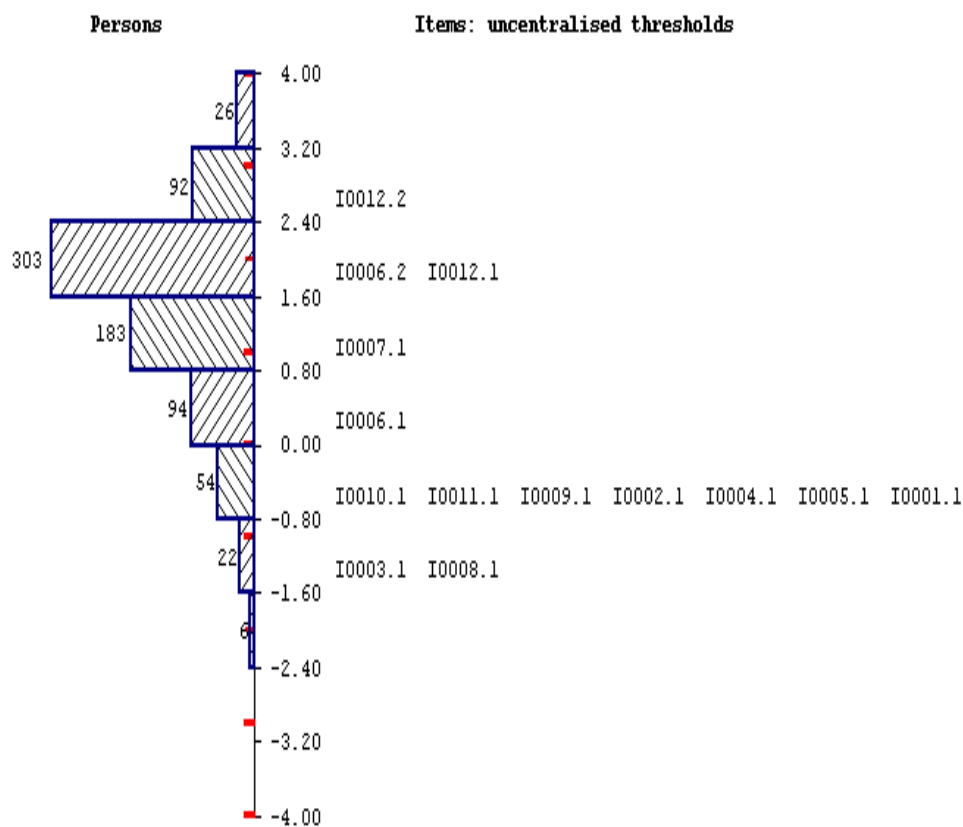


Figure 5.6 Students' Reading Comprehension and Item Thresholds Calibrated on the Same Scale

Notes:

1. Person measures are vertically left
2. Uncentralised thresholds are vertically right. I0012.2 means Item 12, threshold 2, and so on.

Differential Item Functioning (DIF) by Gender

Items 10 and 12 showed no statistically significant main effect for gender, with girls scoring higher ($F=2.56$, $df=6,1$, $p=0.11$) and ($F=5.26$, $df=6,1$, $p=0.02$) respectively, in which, the critical level is $p<0.01$. Item 10 was concerned with the relationship between humans and food in the Food Chain and Item 12 was concerned with from where the passage of writing on Food Chains was taken. While the items were not designed so that answers were item-dependent, it was likely in retrospect that some item dependency occurred because the item content was perhaps too similar. Item 10 and 12 graphs are shown in Figure 5.7 and Figure 5.8. Others, such as items 1, 3 and 6, also showed that girls performed better on the Reading Comprehension test than the boys, but these too were not statistically significant (see, for example, Figure 5.9).

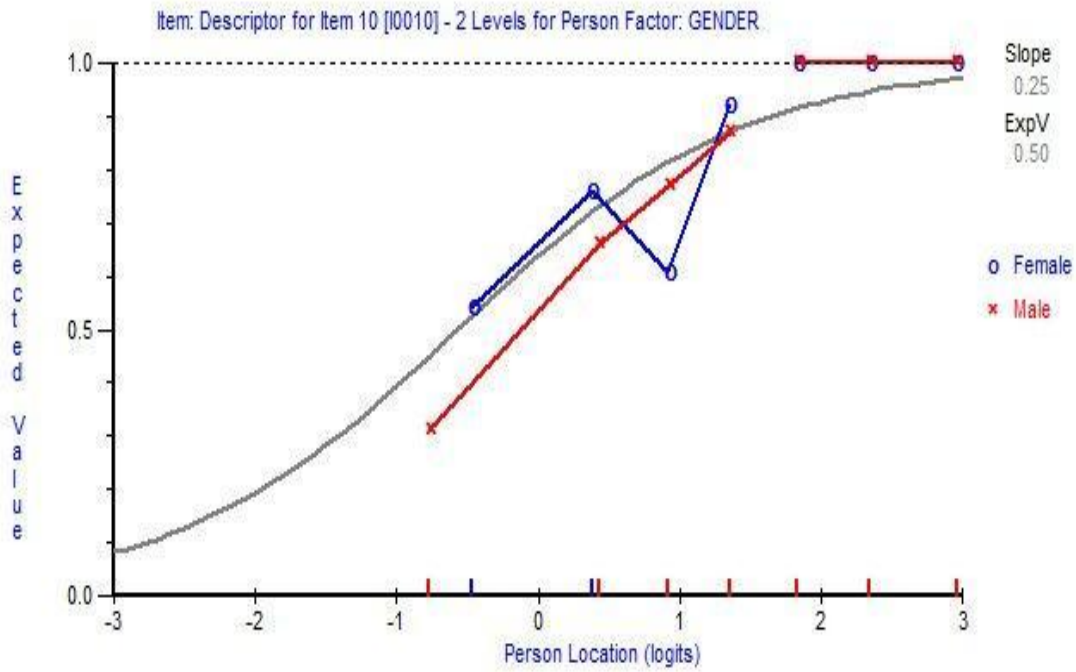


Figure 5.7 Item Characteristic Curves by Gender for Reading Comprehension Item 10

Notes.

1. Broken lines are due to technical errors of Item Characteristic Curves for Reading Comprehension Measure.
2. For the graph: No statistically significant main effect for females, $F=2.56$, $df=6,1$, $p=0.11$.

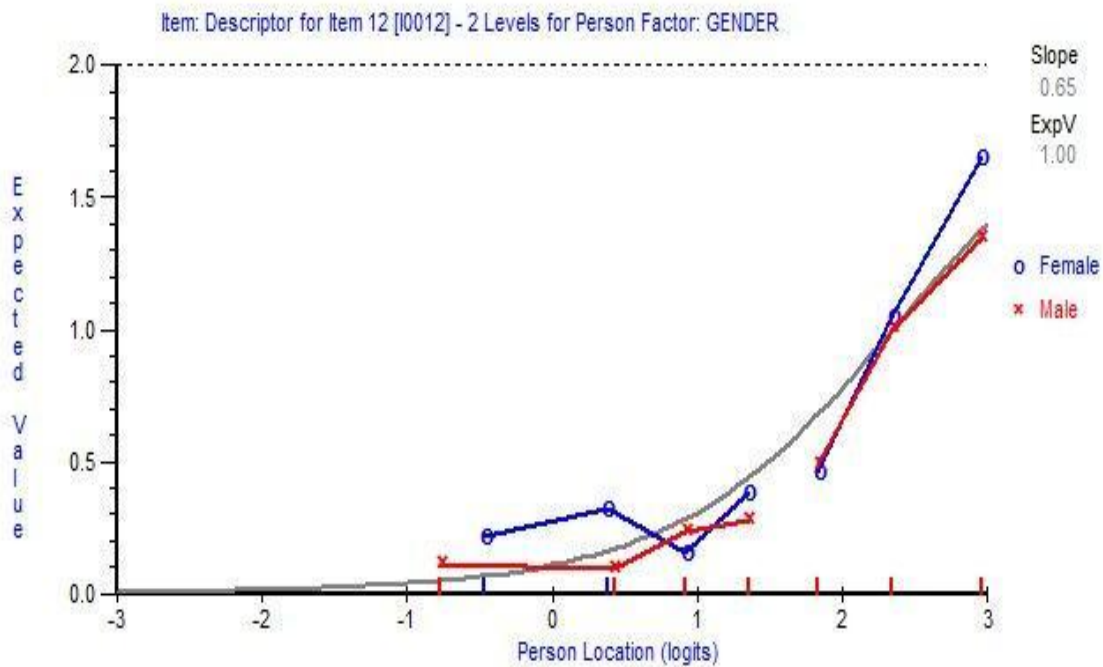


Figure 5.8 Item Characteristic Curves by Gender for Reading Comprehension Item 12

Note. No statistically significant main effect for females, $F=5.26$, $df=6,1$, $p=0.02$

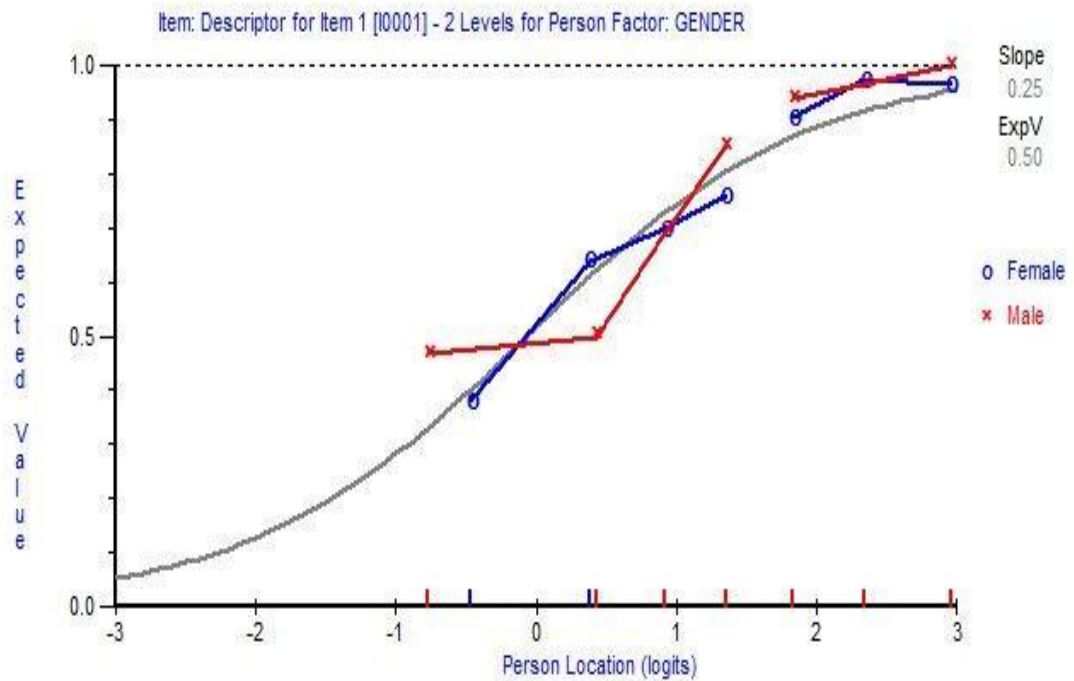


Figure 5.9 Item Characteristic Curves by Gender for Reading Comprehension Item 1

Note. No statistically significant interaction effect, $F=1.05$, $df=6,1$, $p=0.31$

Differential Item Functioning (DIF) by Type of Language Instruction

Items 2, 5, 6, and 9 showed a statistically significant main effect for type of teaching (bilingual or monolingual). For example with item 2, $F=15.37$, $df=6,1$, and $p=0.00009$, with bilingual teaching better; and with item 6, $F=43.45$, $df=5,1$ and $p=0.00000$ with bilingual teaching better. Figure 5.10 shows the Item Characteristic Curves by Type of Language Program (Bilingual/Monolingual) for Reading Comprehension item 2. Other items, such as item 8, showed that bilinguals did better than monolinguals, (see Figure 5.11), but this was not statistically significantly higher ($F=1.17$, $df=6,1$, $p=0.09$).

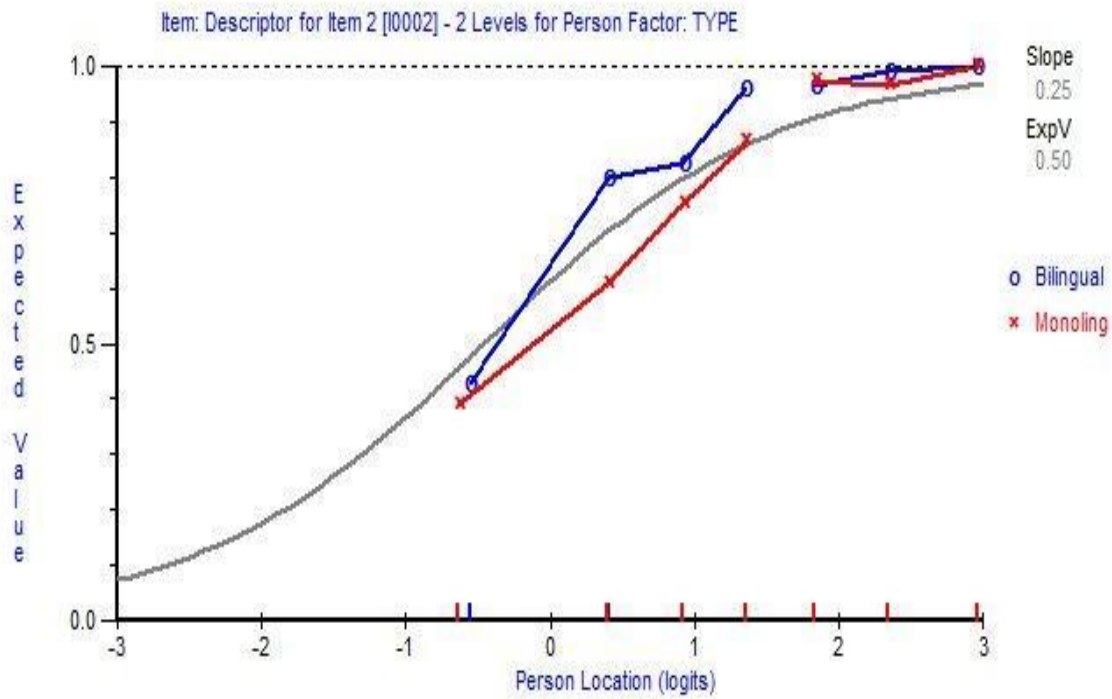


Figure 5.10 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 2

Note. There is a main effect, $F=15.37$, $df=6,1$, $p=0.00009$

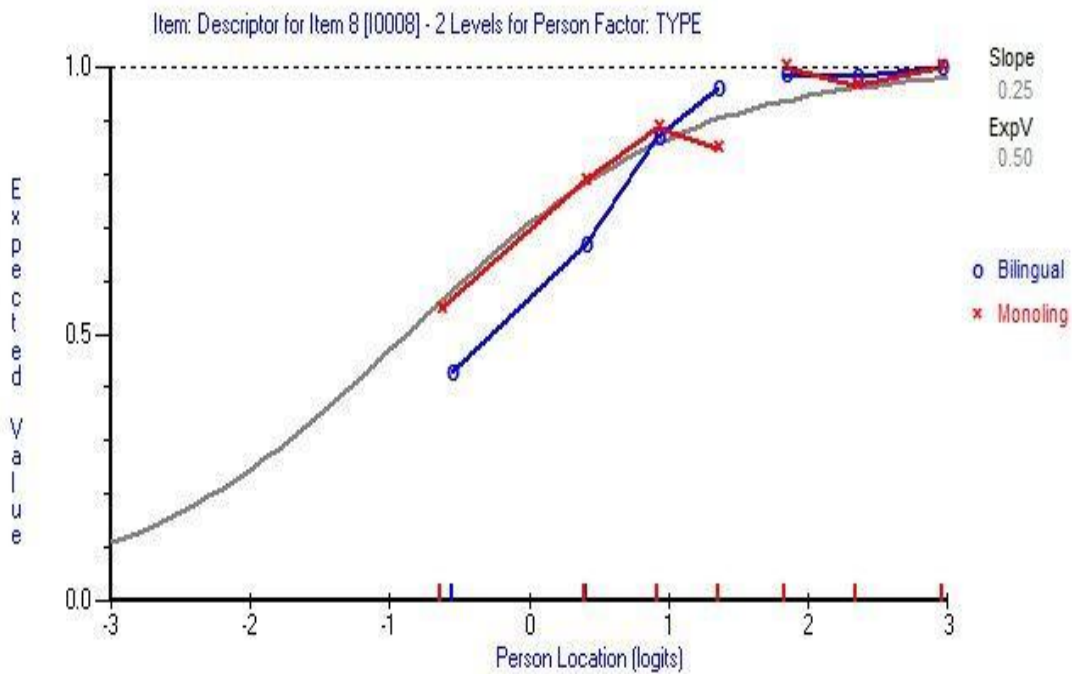


Figure 5.11 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Reading Comprehension Item 8.

Note. No statistically significant main effect, $F=2.90$, $df=6,1$, $p=0.09$

Targeting by Gender and Type of Language Instruction

Overall, girls performed statistically significantly better in the English Reading Comprehension than boys (mean 1.636 logits with SD = 1.05, and mean 1.430 logits with SD = 1.18, respectively) and $F=6.62$, $df=1,778$ and $p=0.01$ (see Figure 5.12). In addition, bilinguals performed statistically significantly better in the English Reading Comprehension than monolinguals (mean 2.094 logits with SD = 0.85, and mean = 0.982 logits with SD = 1.07 respectively) and $F=257$, $df=778,1$ and $p=0.00000$ (see Figure 5.13).

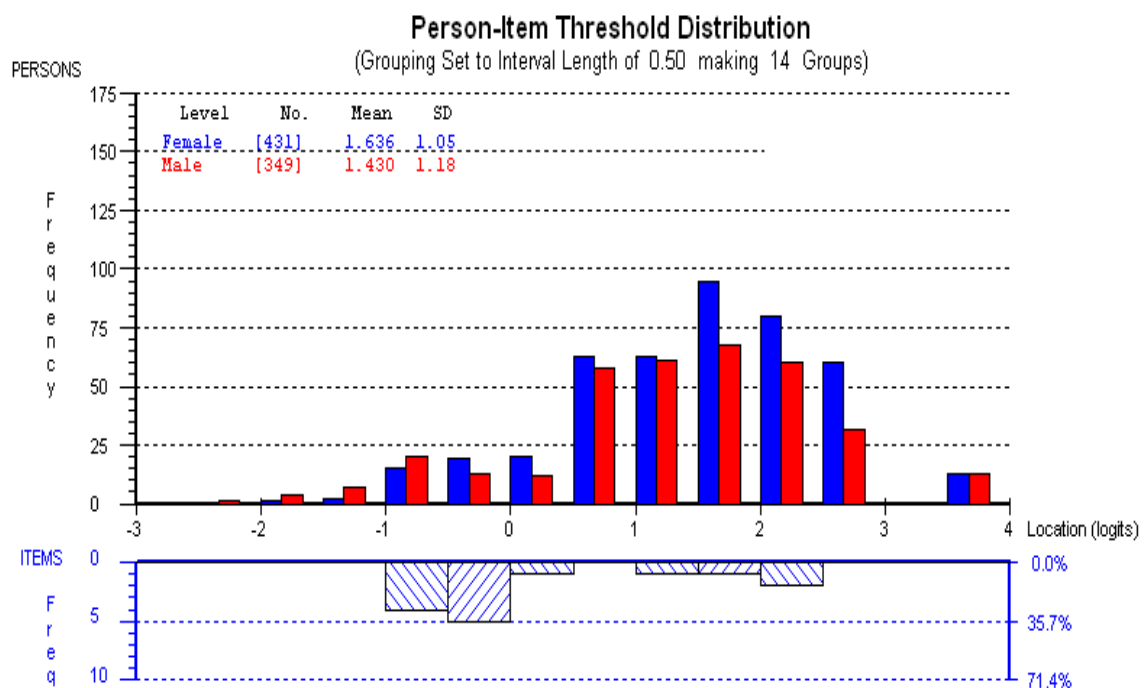


Figure 5.12 Targeting of Reading Comprehension by Gender

Notes:

1. The person measures are on the top-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the bottom-side from easy (LHS) to hard (RHS). $F= 6.62$, $df=1,778$, $p=0.01$, which is statistically significant for females.

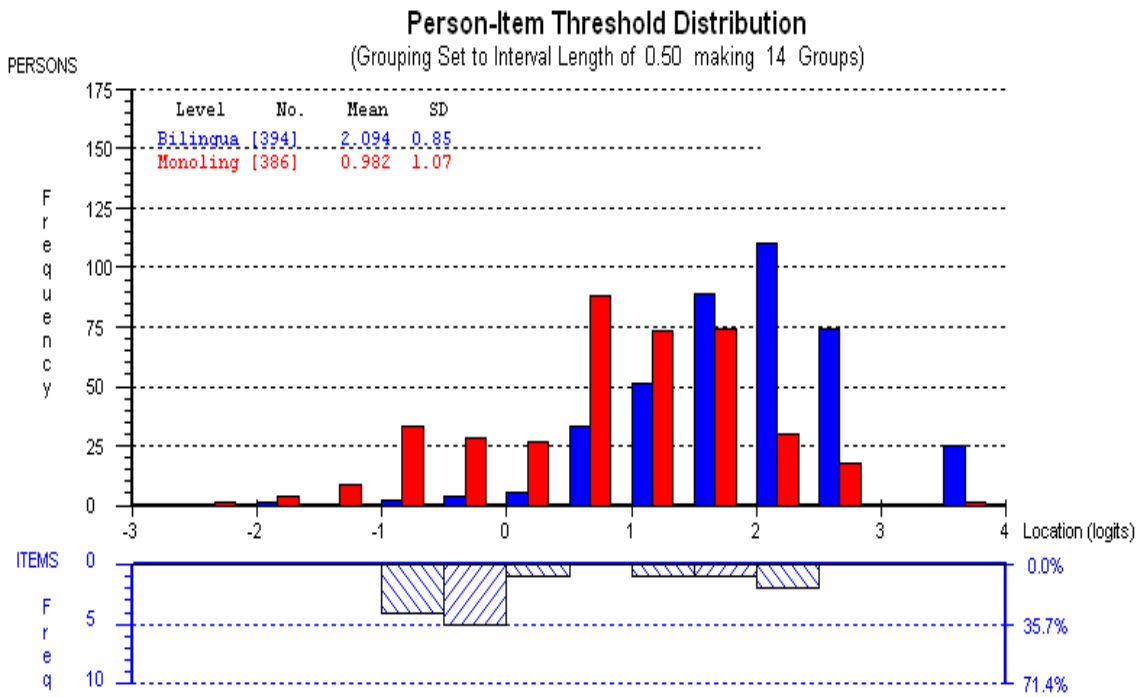


Figure 5.13 Targeting of Reading Comprehension by Type of Language Instruction

Note: $F= 257.89$, $df=1,778$, $p=0.00000$, which is statistically significant for bilingual teaching

Scale of Item Difficulties

Table 5.5
Order of Difficulty of Items on the Linear Scale

Item No.	Item Location	Item Statements on the Food Chain	About
3 (easy)	-0.995	Need for energy	link between plants and animals
8	-0.880	Consumers and decomposers	example of carnivore
10	-0.567	Humans in the food chain	what do humans eat?
11	-0.519	Humans in the food chain	what do we call humans?
9	-0.488	Consumers and decomposers	what decomposers eat?
2	-0.454	Need for energy	where do plants get their energy?
4	-0.379	Sunlight and producers	what do plants use to grow?
5	-0.323	Sunlight and producers	what do we call plants?
1	-0.064	Need for energy	what happens if there is no food?
6	0.985	Sunlight and producers	plant food
7	1.337	Consumers and decomposers	what are primary consumers?
12 (hard)	2.348	Humans in the food chain	why is the food chain important?

The items were ordered from easy to hard on a linear scale (see Table 5.5) so that it can be seen which items are easy and which are hard. The easiest item involved the need for energy and the link between plants and animals, as expected (item 3

difficulty = -0.995 logits). The hardest item involved explaining why the food chain was important, as expected (item 12 difficulty = +2.348 logits).

Discussion and Summary

After review by some teachers and modification, the passage of Reading Comprehension was initially considered appropriate in length for the second language students as again judged by their teachers in Aceh, and the topic was known to the students through previous teaching. Some teachers in Aceh reviewed the Reading Comprehension passage (and the items) and they suggested decreasing the difficulty of the items and the English expression and vocabulary. In doing this, it may be that the items were made too similar.

The original Reading Comprehension test was piloted with sixty students who were either bilinguals or monolinguals. Opinions and suggestions were requested from the students and their teachers. The Reading Comprehension test then was redesigned by taking into consideration the students' and the teachers' opinions. Considered as difficult, the Reading Comprehension test passage was redesigned to contain less unfamiliar vocabulary, and pictures were. This led to the reconstruction of the test with more pictures and repetitive words. Feedback and opinions of students and teachers who preferred more pictures on the reading passage due to its benefit in understanding reading were taken into account. With pictures, the passage became more accessible for the students because the pictures imposed clearer meaning on the content/passage. With repetitive words, the passage became more understandable for the students who appeared to have limited vocabulary. However, by redesigning the test, it seemed that the items were made too similar, which led to the test not having enough difficult items.

It appeared that this redesigned Reading Comprehension passage was not so easy for some students and they could not answer the items as predicted. Some students with a medium-ability could answer the medium items correctly; some had difficulty with the predicted easy items. Similarly, it was predicted that the higher-ability students would be able to answer the easy items and medium difficulty items correctly, but they were unable to do so for all items. This resulted in a less than ideal agreement among the 780 students about the item difficulties along the scale. One reason for the disagreement about the item difficulties was probably guessing. It seemed that some

students still found the items difficult, even though the test had been redesigned prior to the actual test, after the piloting stage, and they guessed the answers to the items that they found difficult.

The presumed guessing that some of the students did is understandable due to their lack of English ability. They had only been learning English as a foreign language for about 30-40 hours in the semester when the test was conducted. Though the curriculum stated that their ability was in line with the Reading Comprehension being tested, much of the time the teaching-learning progress ran much slower than the curriculum suggested. Teachers who were interviewed regarding this matter confessed that they could not run the teaching-learning process, as stated by the curriculum timeline because it would produce even poorer student achievement. The teachers, therefore, needed to pace the students' progress, which meant that some students were behind the curriculum timeline. This could have resulted in students' possessing very limited English when they sat for the Reading Comprehension test for this research. Their lack of ability in English could probably explain the question of why, after redesigning the Reading Comprehension passage test, with more repetitive words and pictures, the students still found the test difficult.

The construct validity of the test was tested by designing the items in ordered patterns of item difficulty which then were compared with their Rasch-measured item difficulties. That is, the Rasch measured item difficulty order given in Table 6.5 was similar to what was predicted when the items were created and designed. The overall fit to the Rasch Measurement Model, as shown by the item-trait interaction chi-square, was $\chi^2=200.4$, $df = 60$ with $p=0.000$, which was not ideal. This meant that the agreement amongst all the students about the item difficulties along the scale was not ideal.

The Cronbach Alpha was 0.53 and low. However, some other measurement aspects were more satisfactory. The thresholds were ordered in line with the scoring categories and the Scoring Category Curves were appropriately ordered with overall measures. The Item Characteristic Curves were satisfactory in most cases, without being ideal, and targeting, while reasonable, would have been improved if there was a wider range in question difficulty. The student measures ranged from -0.8 to +3.9 logits and the item thresholds ranged from -1.2 to +2.6 logits.

Even though the fit to the Rasch Measurement model was not ideal, the Item Characteristic Curves showed no Differential Item Functioning against gender (male versus female) but the Item Characteristic Curves showed Differential Item Functioning against the type of teaching (bilingual versus monolingual teaching). The Differential Item Functioning (DIF) against type of teaching is interpreted as not really being DIF, but as showing that there is a real difference in output by bilingually-taught against monolingual-taught students, with Bilinguals achieving at a higher standard on most items. This is supported by the difference in overall Rasch measures for reading Comprehension by type of teaching. The bilingually-taught mean is 2.094 with $N=394$ and $SD=0.85$ and the monolingually-taught mean is 0.982 with $N=386$ and $SD=1.07$ ($F=257.89$, $df=1,778$, $p=0.00000$). These results showed that the Rasch measure was not ideal and still needed some improvement.

While girls performed significantly higher in English Reading Comprehension than boys overall — in line with the results from many western countries — only two items showed a statistically significant result for girls (item 10 and item 12). Item 10 was a multiple choice and it is hard to see why it should favour girls. Item 12 involved writing and English comprehension which possibly required more thinking and motivation to write to answer the "why" question, and in retrospect, this could be expected to favour girls over boys, not because the item is biased, but because girls in their early secondary school years (12-13 years) were better than boys, in regard to English Reading Comprehension, as shown on items 1, 3, and 6 (see, for example Table 5.9 for Item 1).

In the case of language instruction (bilingual or monolingual instruction), five items showed DIF, with bilingual instruction being superior (items 2, 5, 6, and 9). Items 2, 5, 6, and 9 involved slightly harder reading words and comprehension than the other items and an understanding of their meaning would have been enhanced through bilingual instruction.

By performing DIF against both gender and language instruction, even with a less than ideal fit to the Rasch measurement model, the reason for the DIF could be worked out. Language of instruction would appear to be an important determinant of

performance in English Reading Comprehension and bilingual instruction was related to a superior overall performance compared to monolingual instruction.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for English Reading Comprehension – based on the linear Rasch-created measures.

CHAPTER SIX

DATA ANALYSIS (PART 2)

EXPERIMENTAL RESULTS FOR READING COMPREHENSION

In this chapter, the linear Rasch measures for Reading Comprehension for the bilingually-taught group, known as the experimental group, were compared with similar measures for the monolingually-taught group, known as the control group, in a pretest/posttest, control/experimental group design. In order for the Reading Comprehension measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group to all be on the same scale and comparable, the measures on the pretest have been equated with those on the posttest.

The Mixed Design ANOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant & Tennant, 2007) was conducted to assess the comparison of two different interventions (bilingually-taught program and monolingually-taught program), and of gender on students' scores from the Reading Comprehension Questionnaire for the pretest and posttest. The Two-Way ANCOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant & Tennant, 2007) was employed to test for main and interaction effects with SPSS to compare bilinguals with monolinguals and student gender.

Results on Mixed Design ANOVA (Split-Plot Design ANOVA)

The Mixed Design ANOVA, which is also known as the Split-Plot ANOVA, was conducted to compare the measure of Reading Comprehension Questionnaire of the two language programs (bilingual program and monolingual program) and of gender (female and male). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Reading Comprehension Questionnaire posttest, run following completion of the two months of teaching English as a foreign language. The Reading

Comprehension Questionnaire pretest was conducted prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences.

A mixed between-within-subjects analysis of variance (a.k.a Mixed Design ANOVA or Split-Plot ANOVA) was carried out to assess the difference between two interventions (bilingually-taught program and monolingually-taught program); and the comparison of gender, on participants' Reading Comprehension measures, across two time periods (pretest and posttest).

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity. The result stated that there was no significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.000, $F(1,778) = 0.068$, $p = 0.794$, partial eta squared = 0.000. Wilks' Lambda is a statistical test used in the multivariate analysis of variance to test whether there are differences between the means of identified groups of subjects on a combination of dependent variables. Smaller values of Wilks' lambda indicate greater discriminatory ability of the function (see http://www.blackwellpublishing.com/specialarticles/jcn_9_381.pdf). Partial eta squared (η_p^2) is the default effect size measure in SPSS.

However there was a main effect for time (pretest and posttest), although small, with Wilks' Lambda = 0.985, $F(1,778) = 12.125$, $p = 0.001$, $\eta_p^2 = 0.015$ (small effect) for the Reading Comprehension measure. The main effect comparing the two types of intervention was significant, with $F(1,778) = 369.341$, $p = 0.000$, $\eta_p^2 = 0.322$ (large effect). According to Cohen's effect size guidelines for $\eta_p^2 = 0.322$ were relatively large size effects (Cohen's η_p^2 values: 0.14 = large effect; 0.06 = medium effect; and 0.01 = small effect) (Cohen, 1988; also cited in Burns, 2000). This suggested a real difference in students' ability in English Reading Comprehension with bilingual students achieved better scores than monolingual students (see Figure 6.1 and Table 6.1 and Table 6.2).

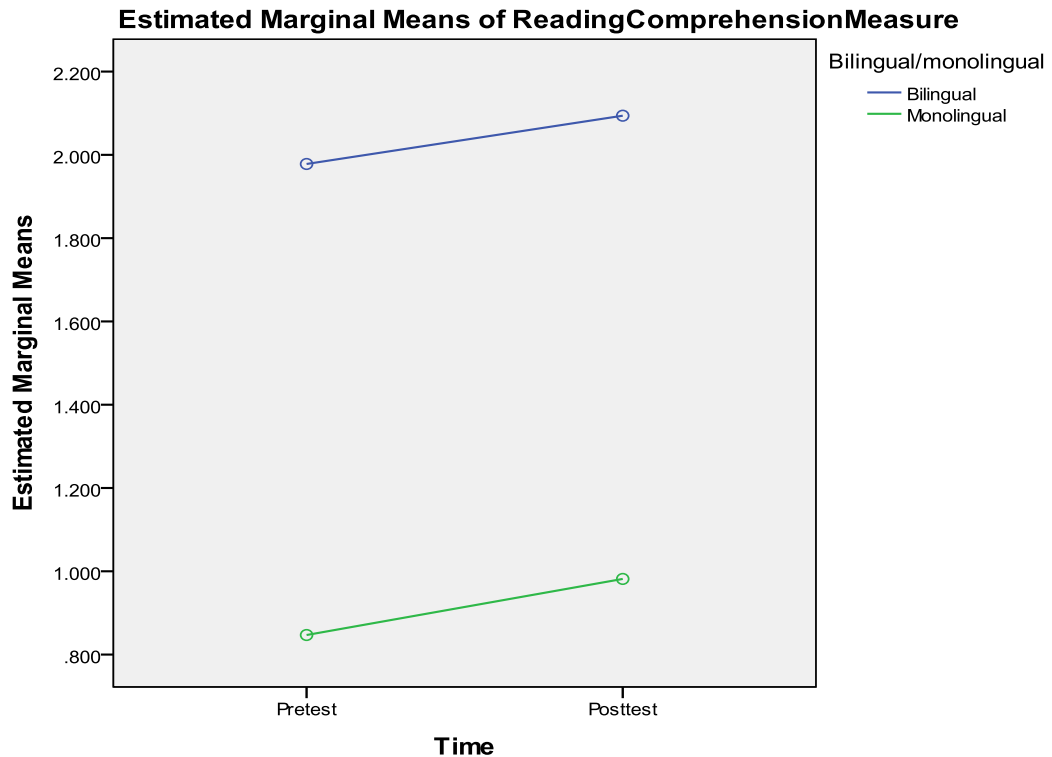


Figure 6.1 Graph of Reading Comprehension by Group

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.999, $F(1,778) = 0.705$, $p = 0.401$, $\eta_p^2 = 0.001$ (a very small effect according to the Cohen's 1988 rules). The main effect for time (pretest and posttest) was not significant either, with Wilks' Lambda = 0.984, $F(1,778) = 12.603$, $p = 0.000$, $\eta_p^2 = 0.016$ (a small effect) for the Reading Comprehension measure. However, the main effect comparing the two types of intervention (control and experimental) was significant, with $F(1,778) = 11.153$, $p < 0.001$, partial eta squared = 0.14 (a large effect size according to Cohen's 1988 regulations). This suggested that there was a real difference in students' ability in English Reading Comprehension, between females and males (see Figure 6.2).

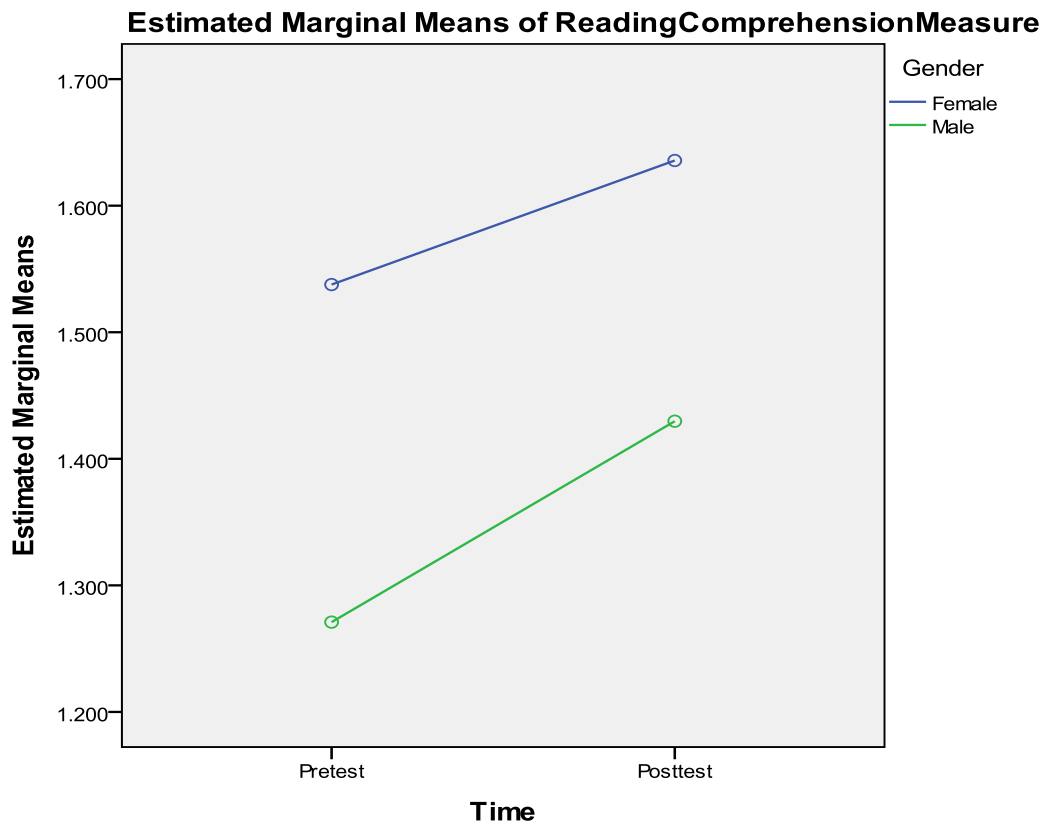


Figure 6.2 Graph of Reading Comprehension by Gender

Table 6.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	1.98	.97	386	.85	.93
Posttest	394	2.09	.85	386	.98	1.07

Table 6.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	1.54	1.03	349	1.27	1.17
Posttest	431	1.64	1.05	349	1.43	1.18

Figure 6.1 and Table 6.1 indicate that bilinguals achieve higher than monolinguals and Figures 6.2, and Table 6.2 indicate that girls have performed moderately better than boys.

Results on Two-Way ANCOVA

The following are the Two-Way ANCOVA results for the Reading Comprehension measures. This test was administered to assess the comparison of the two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught), and gender (male and female). The dependent variable was scores on the posttest, administered following the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language (Time 2). Scores on the pretest administered prior to the start of the programs (Time 1) were used as a covariate to control for individual differences.

Preliminary checks were run to confirm that there were no violations of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for the pretest of Reading Comprehension measure, the result affirmed that there was no significant interaction effect: $F(1, 775) = 0.663, p > .005 (.416)$, with a small effect size ($\eta_p^2 = 0.001$) (see Table 6.3).

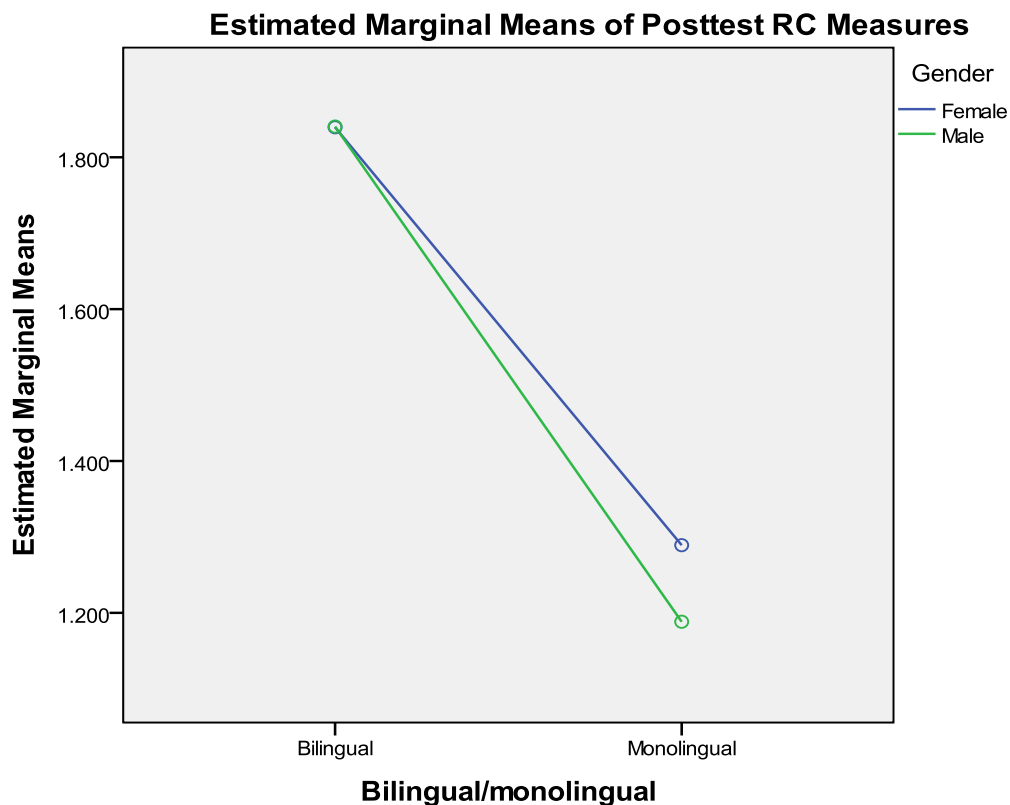
Table 6.3

Interaction Effect and Main Effect of Reading Comprehension (RC) between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest RC Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared (η_p^2)
Corrected Model	389.195 ^a	4	97.299	130.019	.000	.402
Intercept	191.957	1	191.957	256.509	.000	.249
Pretest RC Rasch Measure on Posttest Scale	141.673	1	141.673	189.316	.000	.196
Bilingual monolingual	51.115	1	51.115	68.304	.000	.081
Gender	.473	1	.473	.633	.427	.001
Bilingual monolingual *						
Gender	.496	1	.496	.663	.416	.001
Error	579.965	775	.748			
Total	2827.510	780				
Corrected Total	969.160	779				

a. R Squared = .402 (Adjusted R Squared = .398)

Note: the '*' indicates interaction

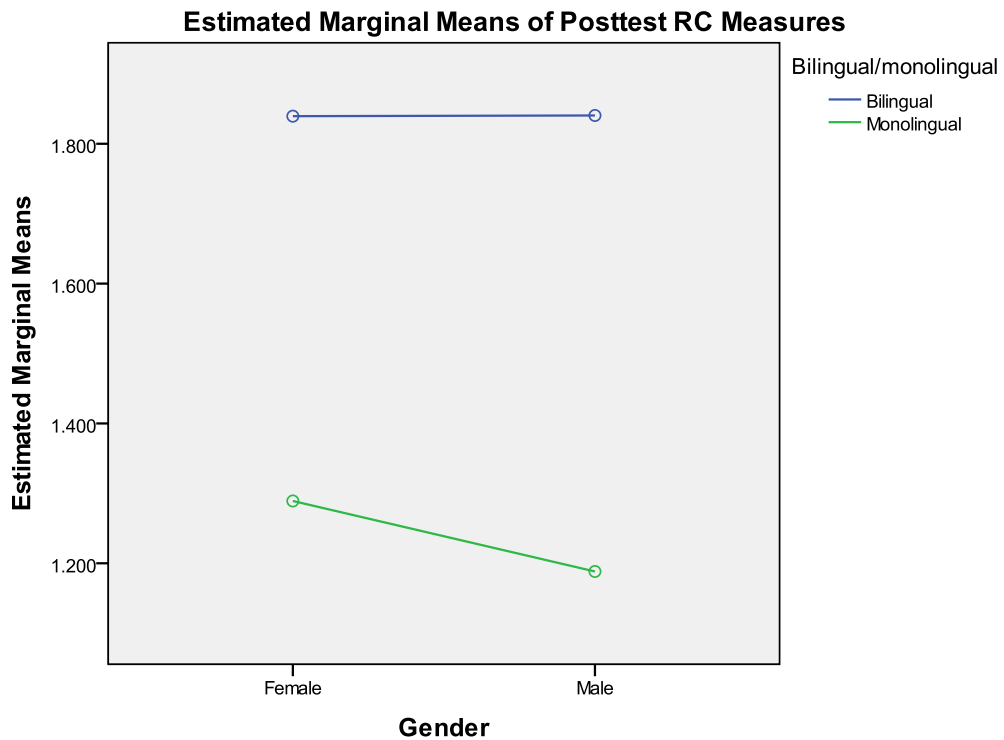


Covariates appearing in the model are evaluated at the following values: Pretest RC Rasch Measure on Posttest Scale = 1.41833

Figure 6.3 Graph of Reading Comprehension by Group (Bilingual/monolingual) with Pretest as Covariate

The main effect for the language program was statistically significant with $F(1,775) = 68.304, p = 0.000$, with $\eta_p^2 = 0.81$ (a large effect size according Cohen's 1988 regulations); while that for gender was not, with $F(1,775) = 0.633, p > 0.005 (0.427)$ with $\eta_p^2 = 0.001$ (a small effect size) (see Figures 6.3 and 6.4).

These results reveal that female students' scores on the Reading Comprehension was equally better than those of male students; and female and male students in the bilingual program responded equally well to the female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest RC Rasch Measure on Posttest Scale = 1.41833

Figure 6.4 Graph of Reading Comprehension by Gender with Pretest as Covariate

Summary

Bilingual (N = 394) and monolingual (N = 386) students were given an English Reading Comprehension pretest and posttest within a two month period (equivalent to

16 lessons), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA using the SPSS computer program. There are two main sets of findings: (1) with regards to the Mixed Design ANOVA, and (2) with regards to the Two-Way ANCOVA.

The main findings for English Reading Comprehension, with regards to the Mixed Design ANOVA, are:

1. There was no significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.000, $F(1,778) = 0.068$, $p = 0.794$, $\eta_p^2 = 0.000$ (small effect).
2. There was a small main effect for time (pretest and posttest), with Wilks' Lambda = 0.985, $F(1,778) = 12.125$, $p = 0.001$, $\eta_p^2 = 0.015$ (small effect) for the Reading Comprehension measure.
3. The main effect comparing the two types of intervention was significant, with $F(1,778) = 369.341$, $p = 0.000$, $\eta_p^2 = 0.322$ (large effect).
4. For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.999, $F(1,778) = 0.705$, $p = 0.401$, $\eta_p^2 = 0.001$ (small effect).
5. The main effect for time (pretest and posttest) was not significant, with Wilks' Lambda = 0.984, $F(1,778) = 12.603$, $p = 0.000$, $\eta_p^2 = 0.016$ (small effect).
6. The main effect comparing the two types of intervention was significant, with $F(1,778) = 11.153$, $p < 0.001$, $\eta_p^2 = 0.14$ (large effect).

The main findings for English Reading Comprehension, with regards to the Two-Way ANCOVA, are:

1. There was no significant interaction effect: $F(1, 775) = 0.663$, $p > .005$ (.416), with $\eta_p^2 = 0.001$ (small effect).
2. The main effect for the language program was statistically significant with $F(1,775) = 68.304$, $p = 0,000$, with $\eta_p^2 = 0.81$ (a large effect size according Cohen's 1988 regulations).

3. The main effect for the gender was not significant with $F(1,775) = 0.633$, $p > 0,005$ (0.427) with $\eta_p^2 = 0.001$ (a small effect).

The findings indicated that female and male bilingual students achieved better scores of Reading Comprehension than female and male monolingual students did. Female students' scores on the Reading Comprehension were better than those of male students.

The next chapter explains the Rasch analysis with the RUMM2030 computer program for the English Writing Test.

CHAPTER SEVEN

DATA ANALYSIS (PART 3)

RASCH MEASUREMENT OF ENGLISH WRITING

This chapter reports the findings of the Rasch analysis of the Writing posttest data.

The Extended Logistic Model of Rasch (Andrich, 1988a) was used to analyse the Writing Test data (I= 12 ‘items’) in order to create a linear measure of students’ English Writing at the posttest stage of the experiment. Each student’s writing was initially marked in four response categories: ‘poor’ (scored 1), ‘adequate’ (scored 2), ‘good’ (scored 3), and ‘excellent’ (scored 4). These three aspects were; (1) paragraph organization (easiest); (2) text conventions (harder); and (3) text quality (hardest). These aspects are called ‘items’ (see Tables 7.1 and 7.2) in order to be compatible with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010). For the rest, the ‘item’ is called item (without brackets).

Initial Analysis (Posttest)

The analysis involved running the data through the RUMM2030 program and checking for non-fitting items and items with poor discrimination, which were deleted, and then re-running the program. After four analyses, six of the original 18 items were rescored (Items 1, 2, 3, 10, 11, and 12); and another six (items 1, 4, 10, 13, 15, and 17) were deleted, leaving 12 items in the final scale: Items 2, 3, 5, 6, 7, 8, 9, 11, 12, 14, 16, 18. The six deleted items did not have a good fit to the measurement model, for two reasons. First, there were overlapping thresholds and poor Scoring Category Curves. Second, the item residuals were too high. Item residuals are differences between the predicted scores, based on the Rasch parameters, and the actual scores on the items. The RUMM program does not tell the researcher why an item does not fit the Rasch Measurement Model but one can infer that many students with the same measures could not agree on the difficulty of the items all along the scale.

A more detailed reason for why some items did not fit the Rasch model was difficult to determine. However, it was assumed that a substantial cause of the misfit was the students' lack of ability in the writing test. The students were in their first semester of the first year in which English was being taught for the first time. They were still weak in English in general and in English writing in particular. The misfit may have been primarily due to their low command of English writing related to the different classroom cultures (Indonesian's way of writing is different from that of English's way) in the two groups (bilingually-taught and monolingually-taught), in a test situation. Thus, it may be that the two groups of students (bilingually-taught and monolingually-taught) did not agree on the difficulty of the items.

The English writing test consisted of two topics. The first topic was 'My *Idul Fitri* Holiday' and the second one was 'My Family'. The difficulties of the items on the two topics was predicted before data collection, based on logic and experience and the 18 items were placed on an ordered scale (see Tables 7.1 and 7.2). Paragraph Organisation were ordered from easy to hard for Topic Sentence, Concluding Sentence and Supporting Sentences); Text Convention was ordered from easy to hard for spelling, punctuation/capitalisation, and grammar); and Text Quality was ordered from easy to hard for readability, style, and text enjoyability (see Appendix D for the scoring rubric).

The three levels of difficulty were ordered from paragraph organisation (considered the easiest), to text conventions (considered to be hard), and text quality (expected to be harder). For Acehnese students learning English writing for the first time in their first semester of Middle Schools, it was considered that their previous knowledge and experience in Indonesian writing has helped them in understanding and ability in English writing. It was expected that the students would have acquired some paragraph organisation from their previous teaching and learning in Indonesian and, therefore, paragraph organisation was expected to be the easiest, because paragraph structure is similar in Bahasa Indonesia and English.

As students who have just started their English lessons, it was expected that they might find text conventions more difficult than paragraph organisation. The students would have had a small exposure to English and they not only lack depth in

vocabulary, but also in spelling, punctuation/capitalisation, and grammar. Thus, text conventions were considered to be harder than the paragraph organisation.

It was expected that students would find text quality the most difficult in their English writing and that they would need substantial English writing experience to do this well. Some students might be able to acquire paragraph organisation in which they are able to produce paragraphs with a topic sentence, some supporting sentences, and finish with a concluding paragraph. They might also be able to make their writing understandable by using the correct vocabulary, spelling, punctuation, capitalisation, and grammar. But, in the early stages, they generally find it difficult to ensure that the sentences are linked together so that the information ‘flows’ nicely for a reader, and that their writing has some style, so that it can be enjoyed by the reader. Therefore, the text quality was considered the most difficult of all.

Each topic (either My *Idul Fitri* Holiday or My Family), has an order of difficulty of ‘easy’, ‘hard’ and, ‘harder’. For ‘Paragraph Organisation’, the ‘main topic’ was considered easy because it involved simple knowledge. The ‘Concluding Sentence’ was considered ‘harder’ because it involved a little more than simple knowledge. The ‘Supporting Sentences’ were considered ‘harder’ as they involved a higher order of thinking which students needed more practice in and knowledge of in order to write around seven or eight correct sentences.

For ‘text conventions’, ‘spelling’ was considered ‘easy’, because it involved common knowledge of the spelling of a noun, verb, adjective or adverb in English. In Indonesia, the spelling of word classes is among the first aspects taught. ‘Punctuation/capitalisation’ was considered ‘hard’ because it involved ability to recognise punctuation/capitalisation which might not be easy for students to do. ‘Grammar’ was considered the hardest compared with ‘spelling’ and ‘punctuation/capitalisation’, because it involved a higher order of ability and knowledge.

For ‘text quality’, ‘readability’ (i.e. clear and neat handwriting) was considered ‘easy’ because it involved a simple ability that students had already achieved. ‘Style’ (i.e. sentence fluency, for example varied length, good flow and rhythm, and varied structure) was considered ‘hard’ because it involved more ability and skills. ‘Text

enjoyability’ was considered the hardest of all under ‘text quality’ because it involved more advanced ability in order to make readers enjoy the writing.

Marks for the writing were allocated from the highest (4) to the lowest (0), based on the English writing scoring rubric (see Table 3.2). Each piece of writing was marked on the 9 items (for each topic) based on the English writing scoring rubric (see Table 3.2). Table 7.1 is an example of the marking used for Writing.

Table 7.1
Writing Marking Example for Topic 1 ‘My Idul Fitri Holiday’

Student									
ID	Item 1 (Title: My Idul Fitri Holiday)								
	Topic. Sent.	Conc. Sents.	Supp. Sents.	Spelling	Punc./ Cap.	Grammar	Read.	Style.	Text Enj.
001	3	1	4	2	2	2	3	3	2
002	1	1	4	3	3	2	3	3	2
003	3	2	4	3	3	2	3	3	3
004	1	2	4	3	3	2	3	3	3
005	3	1	4	2	2	2	3	2	2
006	3	2	4	3	3	2	2	3	2
007	3	2	4	3	3	2	3	3	2
008	3	3	4	3	2	3	3	2	3
009	3	1	4	3	3	2	3	3	2
010	3	1	4	3	3	2	3	2	2

The original behaviour of 18 items (item difficulties for both writing topics) for the difficulties prior to the Rasch analysis and after the Rasch analysis differs. This is due to the fact that the 18 item difficulties after the analysis were fitted to the Rasch model. Table 7.2 shows the original of 18 item difficulties prior to the analysis (with ‘Topic 1 My Idul Fitri Holiday, topic sentence the easiest, and Topic 2 My Family, text enjoyability, the most difficult).

Table 7.2
Initial Marking of English Writing Essay as ‘Items’ (Scored in terms of 18 Items for Compatibility with RUMM2030)

No.	Original Number	
1	1	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Topic Sentence
2	2	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Concluding Sentence
3	3	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Supporting Sentences
4	4	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Spelling
5	5	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Punctuation and Capitalisation
6	6	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Grammar
7	7	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Readability
8	8	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Style
9	9	Topic 1 ‘My <i>Idul Fitri</i> Holiday’ Text Enjoyability
10	10	Topic 2 ‘My Family’ Topic Sentence
11	11	Topic 2 ‘My Family’ Concluding Sentence
12	12	Topic 2 ‘My Family’ Supporting Sentences
13	13	Topic 2 ‘My Family’ Spelling
14	14	Topic 2 ‘My Family’ Punctuation and Capitalisation
15	15	Topic 2 ‘My Family’ Grammar
16	16	Topic 2 ‘My Family’ Readability
17	17	Topic 2 ‘My Family’ Style
18	18	Topic 2 ‘My Family’ Text Enjoyability

Note: The item difficulties for the Writing test have been ordered from Topic 1 ‘My *Idul Fitri* Holiday’ to Topic 2 ‘My Family’, and scoring is set out in terms of ‘items’ for compatibility with the RUMM2030 computer program.

Unlike Table 7.2 that shows the original of 18 item difficulties, before being analysed by RUMM230, Table 7.3 shows the 12 marked items whose data fit the Rasch measurement model, with Topic 1 My *Idul Fitri* Holiday concluding sentence as the easiest marked item and Topic 2 My Family text enjoyability as the most difficult marked item.

Table 7.3

Final Marking of English Writing Essay as Items (Scored in terms of 12 Items for Compatibility with RUMM2030)

No.	Original Number	
1	2 (easiest)	Topic 1 'My <i>Idul Fitri</i> Holiday' Concluding Sentence
2	3	Topic 1 'My <i>Idul Fitri</i> Holiday' Supporting Sentences
3	5	Topic 1 'My <i>Idul Fitri</i> Holiday' Punctuation and Capitalisation
4	6	Topic 1 'My <i>Idul Fitri</i> Holiday' Grammar
5	7	Topic 1 'My <i>Idul Fitri</i> Holiday' Readability
6	8	Topic 1 'My <i>Idul Fitri</i> Holiday' Style
7	9	Topic 1 'My <i>Idul Fitri</i> Holiday' Text Enjoyability
8	11	Topic 2 'My Family' Concluding Sentence
9	12	Topic 2 'My Family' Supporting Sentences
10	14	Topic 2 'My Family' Punctuation and Capitalisation
11	16	Topic 2 'My Family' Readability
12	18 (hardest)	Topic 2 'My Family' Text Enjoyability

Note. The misfitting 'items' are 1, 4, 10, 13, 15, and 17 and they were deleted.

The following the output from the RUMM2030 program supports the creation of a linear scale of Writing English with 12 'items'.

Output from the Final Rasch Analysis (Posttest)

Standardised Fit Residuals

In an ideal measurement scale, the standardised fit residuals should yield an approximately normal distribution with mean value = 0 and standard deviation value = 1, with minimal differences between the actual and expected test score values. The fit residual data for students of the Writing test had a mean near zero ($M = -0.35$) and standard deviation near 1 ($SD = 0.7$) and for items the values were $M = -1.89$ and $SD = 2.1$. This means that the student measure-item response pattern was not ideal and that some revision of the marking scheme is probably needed in any future use of the scoring rubric. Table 7.4 shows item and person overall fit statistics for the Writing test.

Table 7.4
Overall Fit Statistics for the Writing Test (N=780, I=12)

ITEM-PERSON INTERACTION				
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	-1.888	-5.243	-0.346
Standard Deviation	3.404	2.094	3.644	0.701

Notes.

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The standardized fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model. They are satisfactory for the students (persons) but not ideal for the 'items'.

Dimensionality and Item-Trait Interaction

The item-trait interaction chi-square was 487.25, $df=108$, and $p=0.00$ (see Table 7.5). This shows that there was a significant interaction between the students and the scores on the items in the Writing test, indicating that there was not an ideal agreement amongst the students about the item difficulties all along the scale. This means that an accurate prediction of each student's scoring on each item could not be achieved using a single parameter for each student (that is the person measure), and a single parameter for each item (that is the item difficulty), as required for a good fit to the Rasch measurement model. While 'how accurate' one needs is a matter of conjecture, how one could obtain a better fit to the measurement model for English as a second language for Indonesian students should be researched in the future.

A factor analysis of the components of the residuals supports the view that the measure is not unidimensional and that there are at least two factors present. The first eigenvalue is 2.9, and it is above the chance level (see <http://www.rasch.org/rmt191h.htm>). Eigenvalues provide information about how much discriminating ability a function possesses.

Table 7.5
Item-Trait Interaction for the Writing Scale

Total Item Chi-Square	487.25
Separation Index (Reliability)	0.94
Total Degree of Freedom	108.00
Total Chi-Square Probability	0.00
Cronbach Alpha (Reliability)	0.95

Notes

1. The Index of Person Separation (Andrich & Van-Schoubroeck, 1989) is interpreted like a Cronbach Alpha (Cronbach, 1951) and is high and good. Cronbach Alpha is based on the raw scores and the Person Separation Index is based on the Rasch parameters.
2. The item-trait interaction test is a chi-square. The results indicate that the collective agreement amongst the students about the item difficulties along the scale was not ideal.
3. All numbers are given to two decimal points because the errors are only up to two decimal points.

Reliability

The Person Separation Index (PSI) is the standard reliability measure based on the Rasch parameters. It is 0.94 (see Table 7.5 above) and very satisfactory, indicating that there is a reliable separation of student measures in relation to the errors all along the scale. Cronbach Alpha (Cronbach, 1951) the standard reliability measure based on the raw scores, is 0.95 and very satisfactory.

Item Statistics

The RUMM 2030 program calculates the item statistics based on the Rasch parameters. Table 7.6 presents the item difficulties, the standard errors, and the item residuals. While some residuals are satisfactory, others are not ideal, showing that not all items fit the measurement model as well as one would like. It is generally expected that residuals would fall within the range -2.5 to + 2.5 but, as shown in Table 7.6, some residuals were outside this range. Deleting these items and re-analysing the data did not improve the fit to the measurement model.

Table 7.6
Locations, Standard Errors, Residuals and Chi-Squares for Writing ‘Items’

Item	Location	SE	Residual
12	-5.499	0.089	1.361
3	-5.291	0.091	-0.353
16	-1.714	0.101	-1.733
7	-1.662	0.102	-2.508
14	-0.864	0.095	-1.939
5	-0.699	0.098	-4.406
8	0.216	0.097	-3.980
18	0.417	0.101	-2.808
9	2.384	0.103	-3.704
6	2.940	0.115	-4.206
11	4.824	0.172	1.078
2	4.947	0.187	0.541

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.

Scoring Category Curve

The RUMM 2030 program produces Scoring Category Curves that examine scoring reliability and show the relationship between the probabilities of scoring in each category as the measure increases. Each item was originally scored in four categories: poor (scored 0), adequate (scored 1), good (scored 2) and excellent (scored 4). For the Writing test, some items were rescored and were marked with only three scoring categories, while the rest were marked with four scoring categories. This was to ensure a high reliability and consistency of marking. The Scoring Category Curves should show a consistent relationship between the probability of scoring in a particular category and the measure, indicating that the scoring was done consistently and logically.

A Scoring Category Curve for item 11 (with three scoring categories) is shown in Figure 7.1. There is a high probability of scoring category 0 and a low probability of scoring category 1, at the lowest student measures. The probability of scoring category 0 decreases and the probability of scoring category 1 increases, as the student measures increase. The probability of scoring category 1 decreases and the probability of scoring category 2 increases, at the highest student measures. The Scoring Category Curves for

the other items were checked and they showed a logical and consistent scoring as well, but they are not shown here.

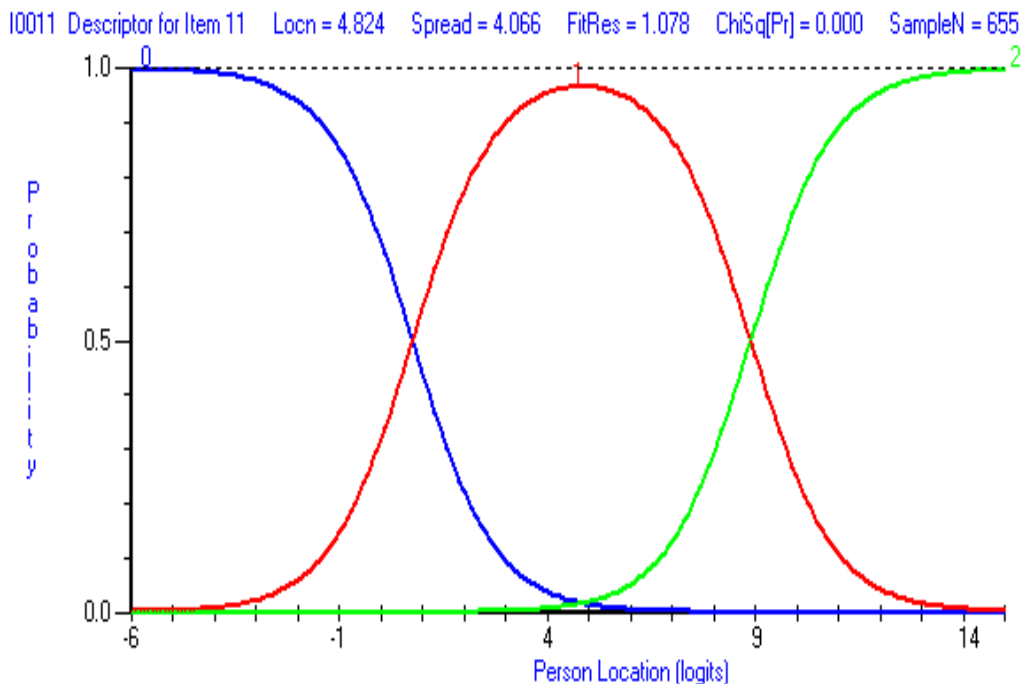


Figure 7.1 Curve for Item Category for Item 11

Note: The blue curve is for a score of 0, the red curve for a score of 1 and the green curve for a score of 2.

Item Characteristic Curves

The Item Characteristic Curve for item 3 (good fitting item) of the Writing test is shown on Figure 7.2. The ogive curve indicates the expected score for the English Writing groups, ranging from the lowest to the highest ability groups. The dots indicate the observed mean measures for the ten student ability groups. When the observed scores closely follow the curve of expected values, the groups are performing as expected on the item. It shows that the item is discriminating well and that as students with higher measures answer the item, they have higher expected values.

The Item Characteristic Curve for item 9 (a not ideal fitting item) of the Writing test is shown on Figure 7.3. The Item Characteristic Curves for the other items were checked. While some were not ideal, they were found to be satisfactory.

10003 Descriptor for Item 3 Locn = -5.291 Spread = 1.814 FitRes = -0.353 ChiSq[Pr] = 0.139 SampleN = 655

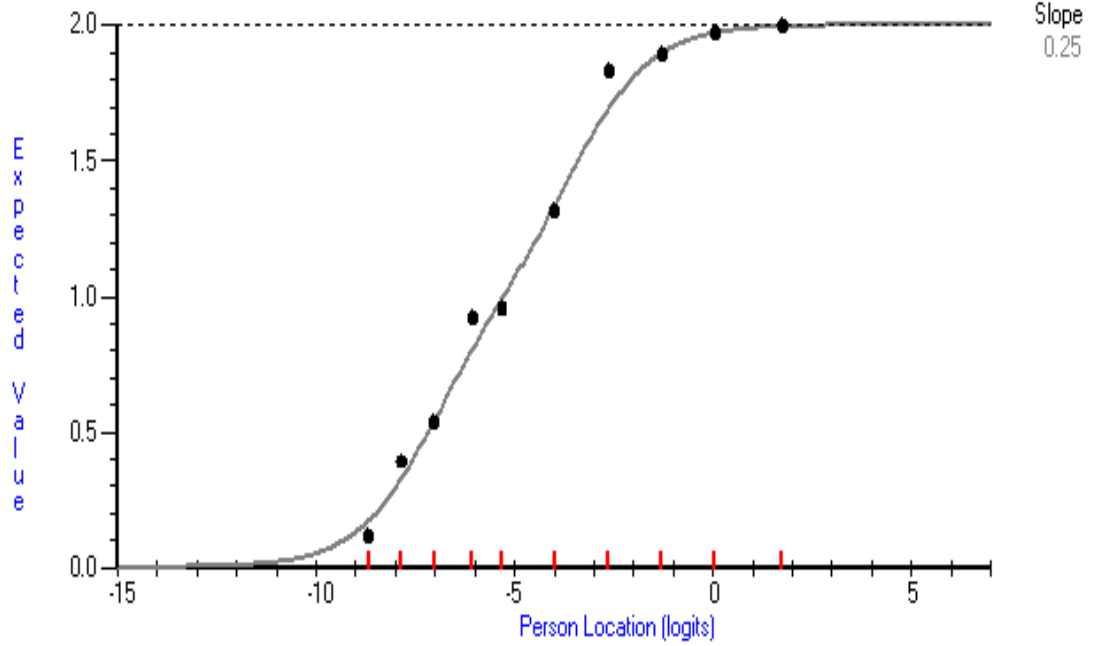


Figure 7.2 Item Characteristic Curve for Item 3

10009 Descriptor for Item 9 Locn = 2.384 Spread = 3.876 FitRes = -3.704 ChiSq[Pr] = 0.000 SampleN = 655

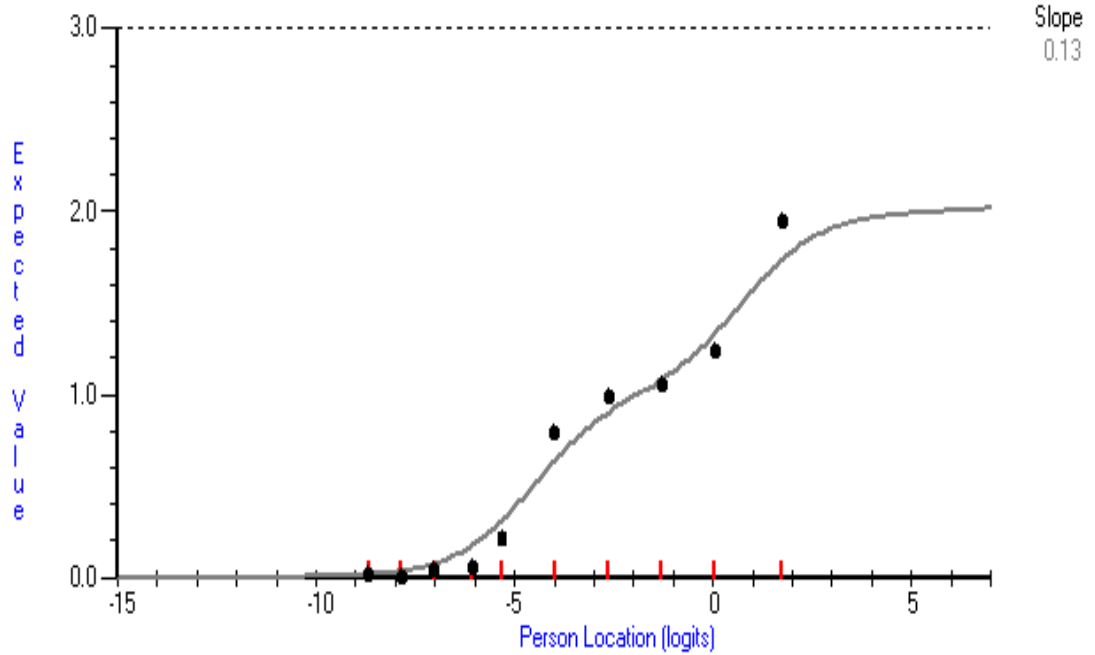


Figure 7.3 Item Characteristic Curve for Item 9 (Not Ideal but Satisfactory)

Person-Item Threshold Distribution (Targeting)

A Person-Item graph (see Figure 7.4) shows the student measures (on the upper side of the scale) and the item difficulties (on the lower side of the scale) from easy (on the left) to hard (on the right). The graph shows that the student measures ranged from -10.5 logits (lowest) to the +6.5 logits (highest), and the item difficulties ranged from -8.5 logits (easiest) to the +10.5 logits (hardest). Measures are considered well-targeted when the range of item thresholds and the range of student measures are about the same. This means that the items were not too hard and not too easy for the students. For this measure, the range of the student measures covered a good range of item difficulties, and so the targeting was very satisfactory. However, some easier items could to be added to the scale to improve targeting and some of the most difficult 'items' could be deleted for these students in any future use of the scale.

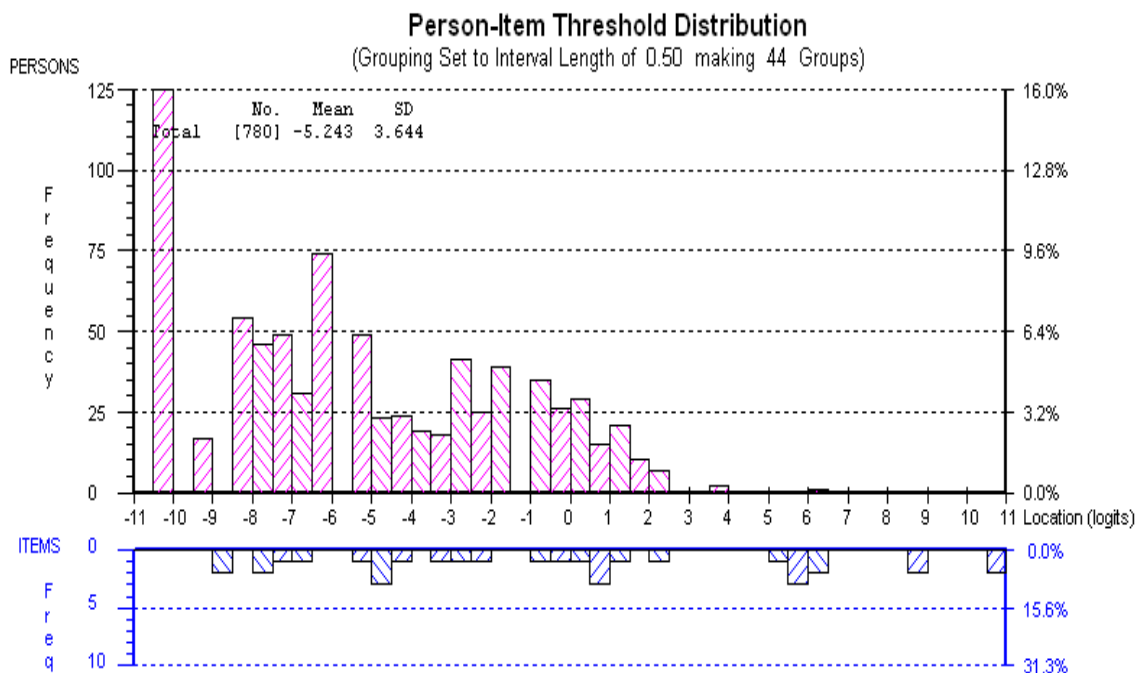


Figure 7.4 Targeting Graph: Student Measures in English Writing (upper-side) and Item Difficulties (lower-side) Calibrated on the Same Scale

Figure 7.4 shows the student measures and the item threshold locations on the same linear scale. Item thresholds are grouped by their locations (difficulties), for example, the locations of 18.2, 9.2, 11.1, and 2.1 are from 0.50 logits to 2.00 logits. The

columns (pink) on the upper-side refer to student measures, which are placed on the same linear scale as the item threshold difficulties, which are placed on the lower-side.

Item thresholds are points between the scoring categories where there is an equal chance of having scores in adjacent categories. In good, consistent measures, the thresholds for the items should be ordered in line with the ordering of the scoring categories, as was the case with this measure (see Figure 7.5).

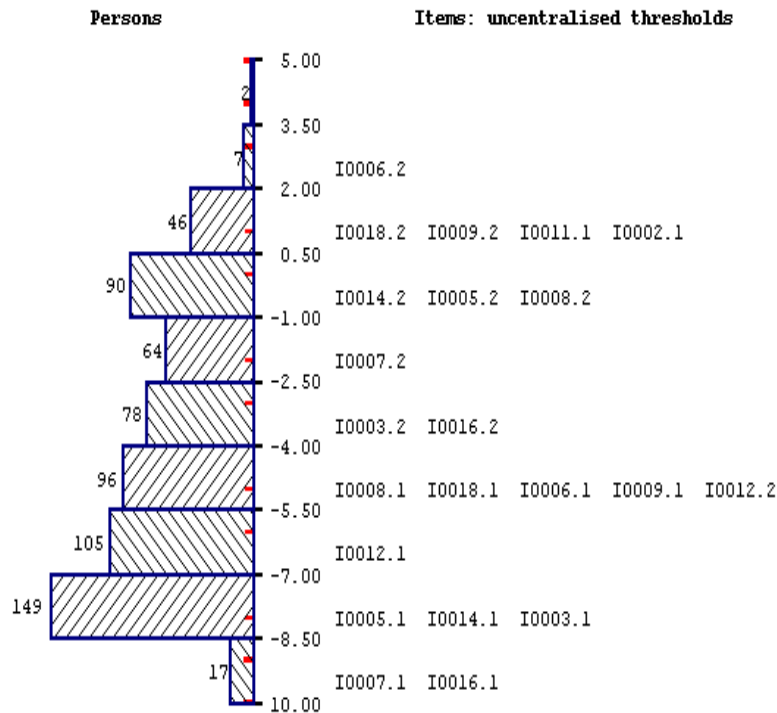


Figure 7.5 Writing Map

Note: I0018.2 means threshold 2 for item 18; I0007.1 means threshold 1 for item 7, and so on.

DIF by Gender

Each of the 12 items of the Writing test showed no statistically significant Differential Item Functioning (DIF) by gender. Item 12 and Item 3 are shown as examples (see Figures 7.6 and 7.7).

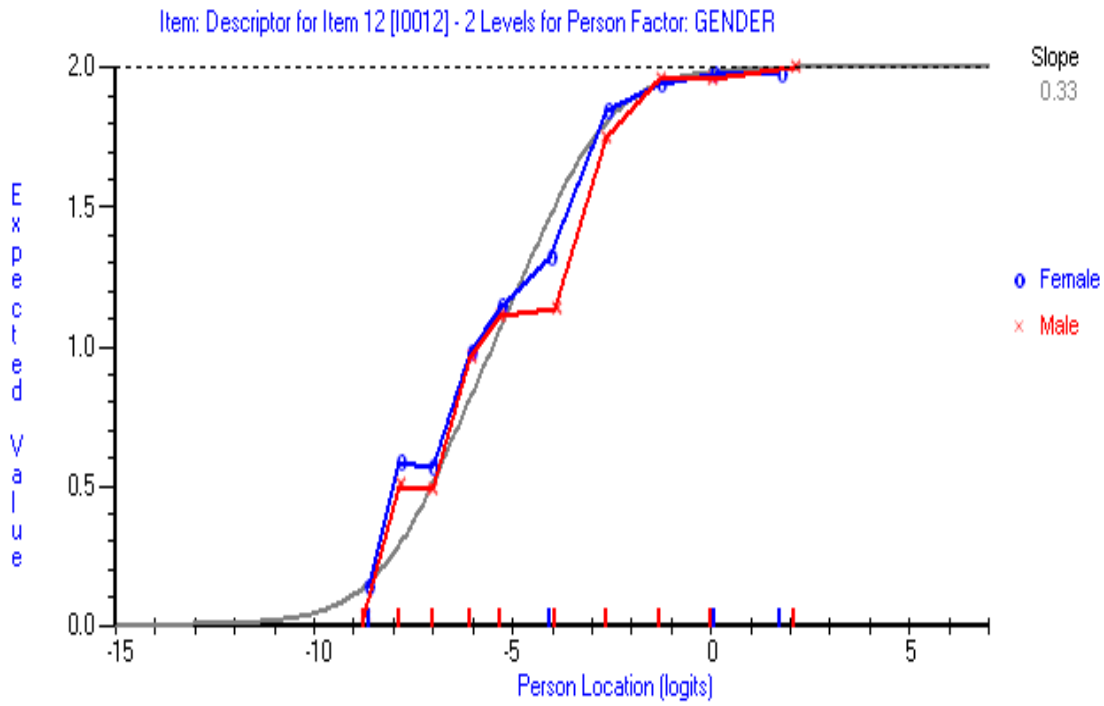


Figure 7.6 Item Characteristic Curve by Gender for Writing Item 12

Note: No statistically significant main effect by gender (no gender bias), $F=1.20$, $df=19,1$, $p=0.27$.

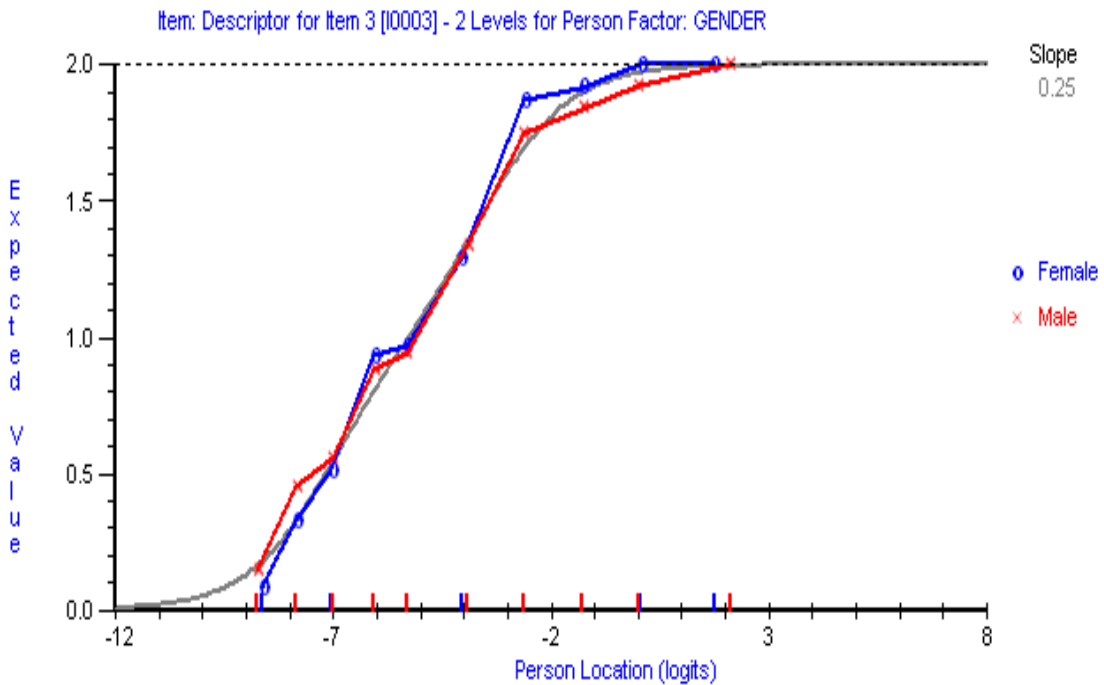


Figure 7.7 Item Characteristic Curve by Gender for Writing Item 3

Note: No statistically significant main effect by gender, no gender bias, $F=1.75$, $df=1,19$, $p=0.18$.

All the Item Characteristic Curves for the other items were checked and showed no statistically significant bias for gender.

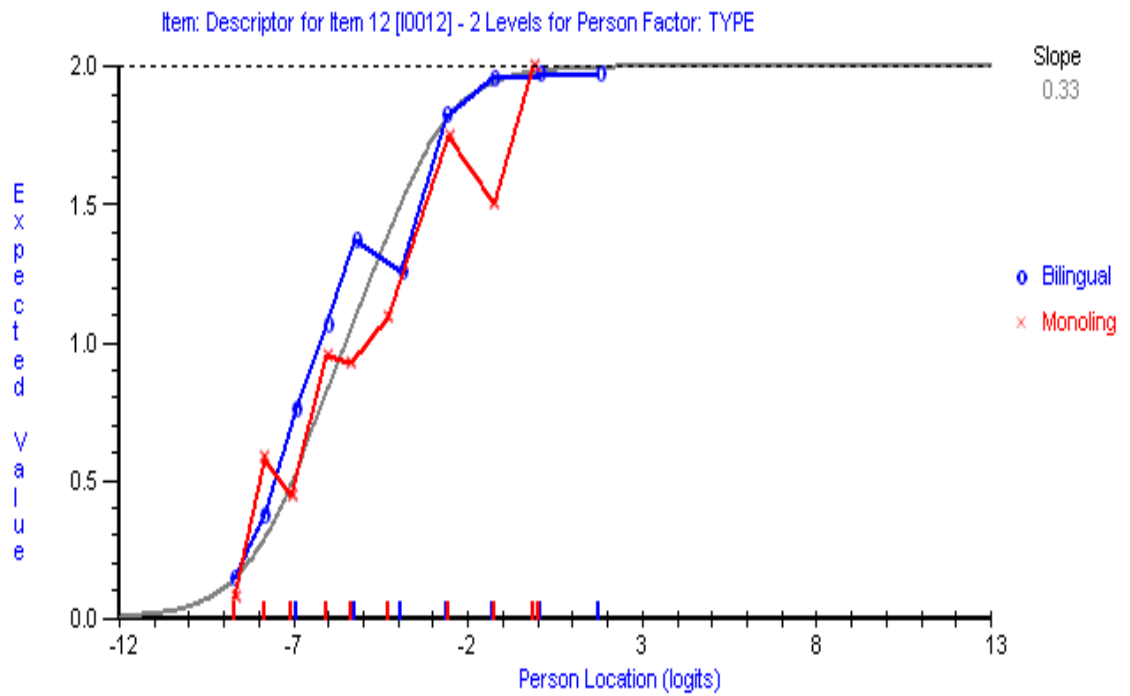


Figure 7.8 Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 12.

Note: No statistically significant main effect by Type of Instruction (no type of language instruction bias), $F=0.004$, $df=1,18$, $p=0.95$

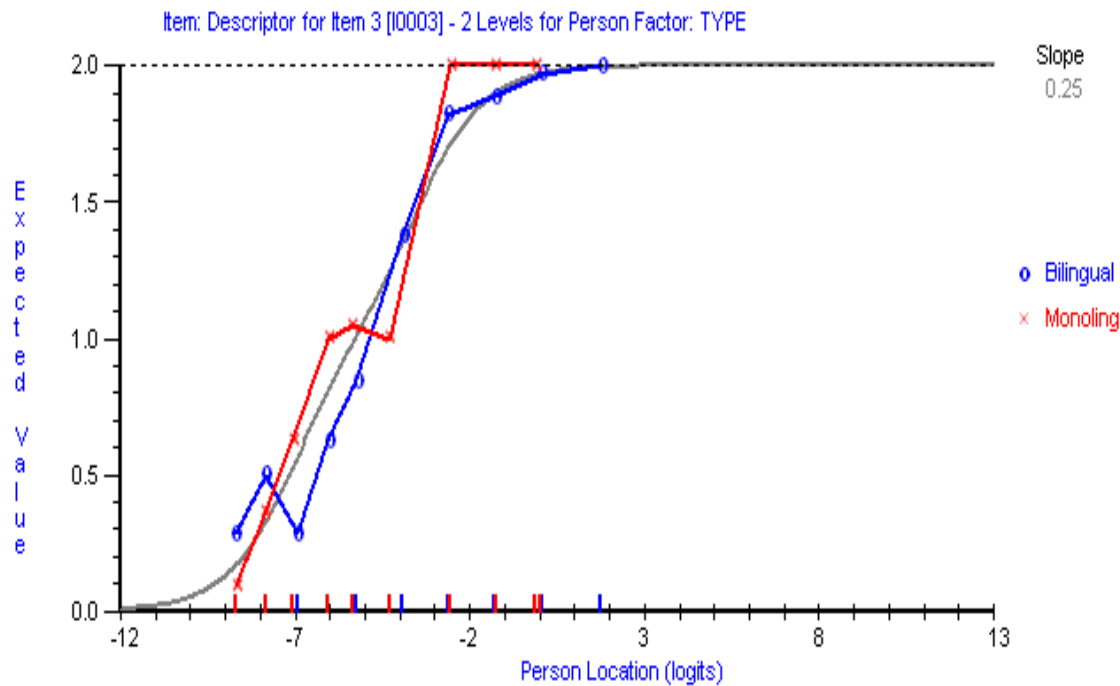


Figure 7.9 Item Characteristic Curve by Type of Teaching Methods (Bilingual vs. Monolingual) for Writing Item 3

Note: No statistically significant main effect by Type of Instruction (no type of language instruction bias), $F=1.67$, $df=1,18$, $p=0.20$

DIF by Type of Language Instruction

Each of the 12 items of the Writing test showed no statistically significant Differential Item Functioning (DIF) by Type of Language Instruction (bilingual versus monolingual). Item 12 and Item 3 were taken as examples (see Figures 7.8 and 7.9). This shows that there was no inherent bias in the items with regard to either of the language instruction types.

Targeting by Gender and Type of Language Instruction

Girls were statistically significantly better in Writing than boys ($F= 59.39$, $df=1,778$, $p=0.00000$, see Figure 7.10) and bilinguals did statistically significantly better in Writing than students in monolingual schools ($F= 706.45$, $df=1,778$, $p=0.00000$, see Figure 7.11).

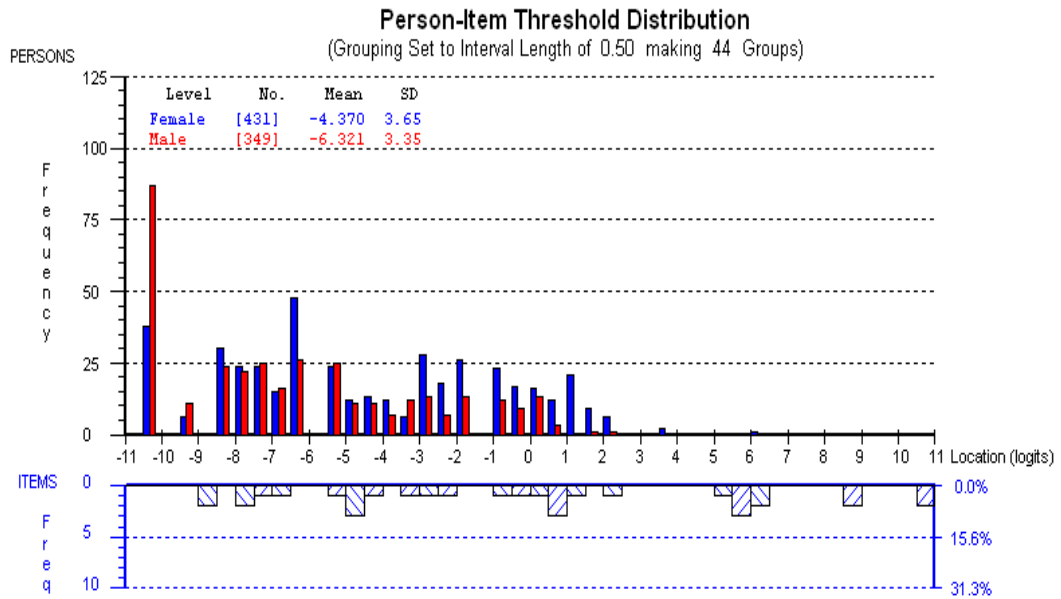


Figure 7.10 Targeting of Writing by Gender

Notes:

1. The person measures are on the upper-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the lower-side of the graph from easy (LHS) to hard (RHS). $F=59.39$, $df=1,778$, $p=0.00000$, meaning that the difference between females and males is statistically significant. Females performed better than males.

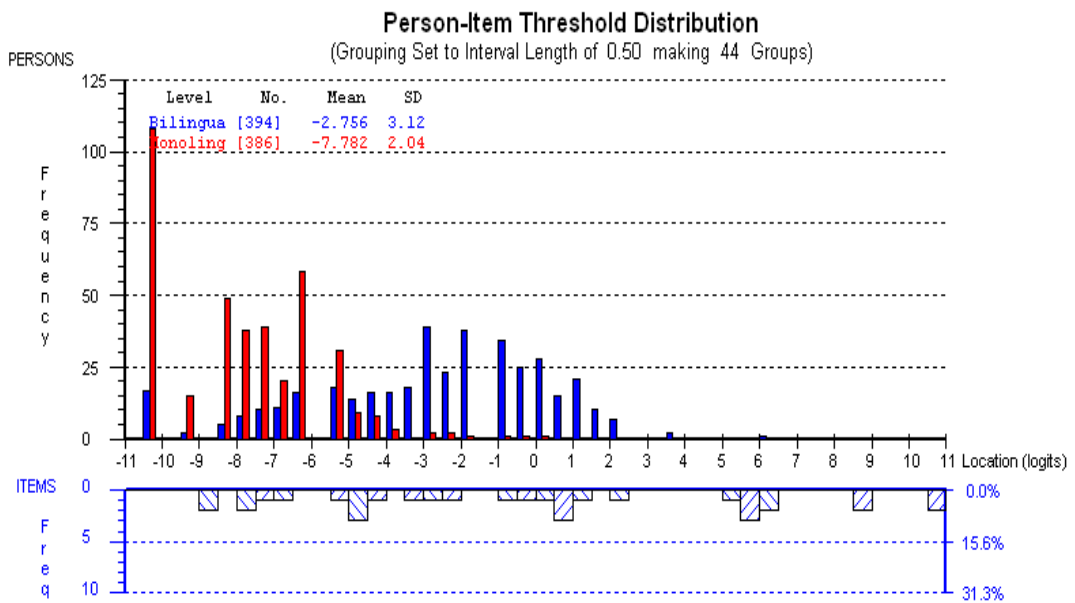


Figure 7.11 Targeting of Writing by Type of Language Instruction

Note: $F=706.45$, $df=1,778$, $p=0.00000$, meaning that the difference between bilinguals and monolinguals is statistically significant. bilinguals performed better than monolinguals.

Scale of Item Difficulties

The corresponding item difficulties for both topics were very similar, showing support for the scoring method and the item difficulty prediction. Table 7.7 shows the items for the Writing Test which have been ordered by difficulty from the easiest to the most difficult on the linear Rasch-created scale. The ordering of the items by difficulty is consistent with the initially predicted conceptualized order, giving support for the construct validity of the scale.

Table 7.7
Ordered Difficulty of Items on the Linear Scale

No	Item Location	Item Statements on the Writing
Easy		
12	-5.50	Topic 2 My Family Supporting Sentences
3	-5.29	Topic 1 My Idul Fitri Holiday Supporting Sentences
16	-1.71	Topic 2 My Family Readability
7	-1.66	Topic 1 My Idul Fitri Holiday Readability
14	-0.86	Topic 2 My Family Punctuation/Capitalisation
5	-0.70	Topic 1 My Idul Fitri Holiday Punctuation/Capitalisation
8	0.22	Topic 1 My Idul Fitri Holiday Style
18	0.42	Topic 2 My Family Text Enjoyability
9	2.38	Topic 1 My Idul Fitri Holiday Text Enjoyability
6	2.94	Topic 1 My Idul Fitri Holiday Grammar
11	4.82	Topic 2 My Family Concluding Sentence
2	4.95	Topic 1 My Idul Fitri Holiday Concluding Sentence
Hard		

Note: Item difficulties (locations) are measured in logits, the log odds of answering successfully.

The items were ordered from easy to hard on a linear scale (see Table 7.7) so that it can be seen which items are easy and which are hard. The easiest item was on Topic 2 My Family Supporting Sentences and the hardest items was on Topic 1 My *Idul Fitri* Holiday Concluding Sentence.

Pretest

The pretest Writing data with 12 ‘items’ were also analysed with the RUMM2030 computer program, in a similar way to posttest Writing data. The results were similar for both pretest and posttest data and, to avoid repetition, the pretest Writing data output is not reported here. The Person Separation Reliability (based on

Rasch parameters) for the pretest Writing data was 0.91 and satisfactory. The Cronbach Alpha reliability for the pretest Writing data (based on the raw marking scores) for the pretest Writing data was 0.93 and also satisfactory.

Discussion and Summary

This chapter shows the results of the data analysis for the Writing test as part of the process of teaching and learning English as a second language in Aceh Province. Out of the original 18 'items', six items showed misfit to the Rasch measurement model and were deleted, leaving 12 'items' to create a linear measure. The Fit Residual data was partially satisfactory, showing some reasonable item-person response patterns, but it was not ideal. The Item-Trait Interaction and principal components analysis of the residuals indicated that two factors were present in the data, that is, the eigenvalue was 2.9, and was therefore above the chance level. However, the Person Separation Index of Reliability and the Cronbach Alpha Reliability were very satisfactory, indicating that there was good separation of measures in comparison to errors. The threshold values and the Scoring Category Curves showed that the scoring categories were used consistently and logically. The Item Characteristic Curves showed good discrimination. So a reasonable linear scale was created which was then used for the ANOVA and ANCOVA analyses which are presented in the next chapter.

The Item Characteristic Curves showed that none of the 12 items exhibited any statistically significant differential item functioning (DIF) by gender and by type of language of instruction (bilingual and monolingual). However, over the Rasch measure for all 12 items together, bilinguals had statistically significantly better Writing ability than monolinguals and girls had statistically significantly better Writing measures than boys, the latter being true even in the pretest, Monolingual Writing data.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for English Writing – based on the linear Rasch-created measures.

CHAPTER EIGHT

DATA ANALYSIS (PART 4)

EXPERIMENTAL RESULTS FOR ENGLISH WRITING

This chapter presents a comparison between the linear Rasch measures for the Writing Test for the bilingually-taught group, also known as the experimental group, and similar measures for the monolingually-taught group, also known as the control group, in a pretest/posttest, control/experimental group design. The measures on the pretest have been equated with those on the posttest in order that the Writing Test measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group are all on the same scale and thus comparable.

A Mixed Design ANOVA, also known as the Split-Plot ANOVA, was employed with the computer program SPSS (Statistical Package for the Social Sciences, Pallant, 2007) to compare the measures of Writing between the two language programs (bilingual program and monolingual program) and for gender (female and male students). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Writing posttest, administered following the completion of the two months of teaching English as a foreign language. The Writing pretest was conducted prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences. Tests for main and interaction effects using the pretest measures as the covariate were used. In order to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity (equality of variances of differences between all combinations of groups), preliminary checks were conducted and found to be satisfactory.

Results for Mixed Design ANOVA (Split-Plot Design ANOVA)

The results showed that there was a significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 0.818, $F(1,778) = 173.41$, $p = 0.000$, partial eta squared (η_p^2) = 0.182. According to Cohen's effect size guidelines, 0.921 is a large effect, as well as 0.481 (Cohen's η_p^2 values: 0.14 = large effect; 0.06 = medium effect; and 0.01 = small effect) (Cohen, 1988; also cited in Burns, 2000). Similarly, there was a significant main effect for time (pretest and posttest), with Wilks' Lambda = 0.79, $F(1,778) = 9089.87$, $p = 0.000$, $\eta_p^2 = 0.921$ for the Writing Test. The main effect comparing the two types of intervention (bilinguals and monolinguals) was also significant, with $F(1,778) = 720.70$, $p = 0.000$, $\eta_p^2 = 0.481$. This suggested a real difference in students' English Writing where bilingual students were able to write English Writing better than monolingual students (see Figure 8.1 and Table 8.1).

It is interesting to note that the posttest measures are lower, on average, than the pretest measures. One explanation for this is that the two month time interval between pretest and posttest measures is insufficient for young Indonesian students to strongly improve their English writing skills. Another explanation is that there was an experimental student effect at pretest but not at posttest. It is possible that a combination of these two effects is responsible for the results.

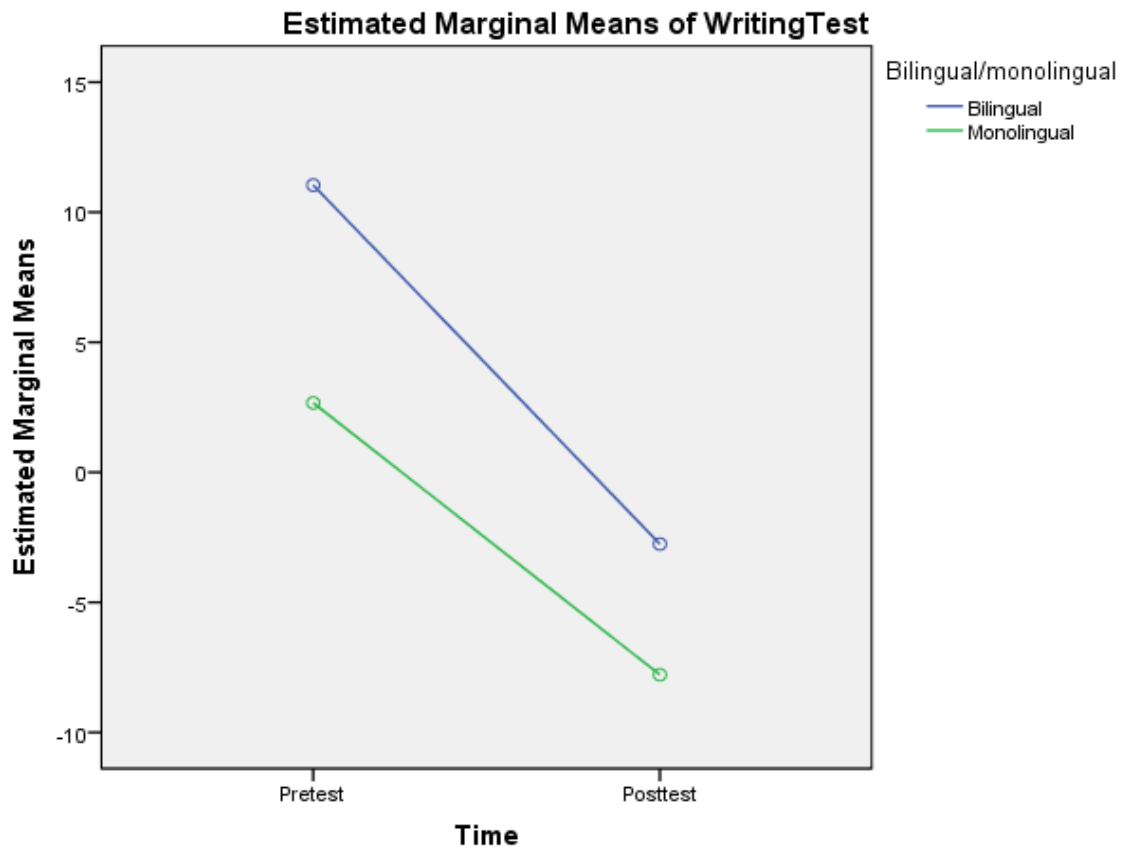


Figure 8.1 Graph of Writing Test by Group

Table 8.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	11.04	5.88	386	2.67	3.53
Posttest	394	-2.76	3.12	386	-7.78	2.04

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.991, $F(1,778) = 7.378$, $p = 0.007$, $\eta_p^2 = 0.009$ (a very small effect according Cohen's 1988 rules). On the other hand, the main effect for time (pretest and posttest) was significant, with Wilks' Lambda = 0.95, $F(1,778) = 7393.32$, $p = 0.000$,

$\eta_p^2 = 0.905$ (a large effect size according to Cohen, 1988) for the Writing Test.

Likewise, the main effect comparing the two types of intervention (monolingually-taught and bilingually-taught) was significant, with $F(1,778) = 47.614, p = 0.000, \eta_p^2 = 0.58$ (a medium effect size according to Cohen's 1988 regulations). This suggested that there was no real distinction in students' ability, between females and males (see Figure 8.2 for plot and Table 8.2 for pretest and posttest measures).

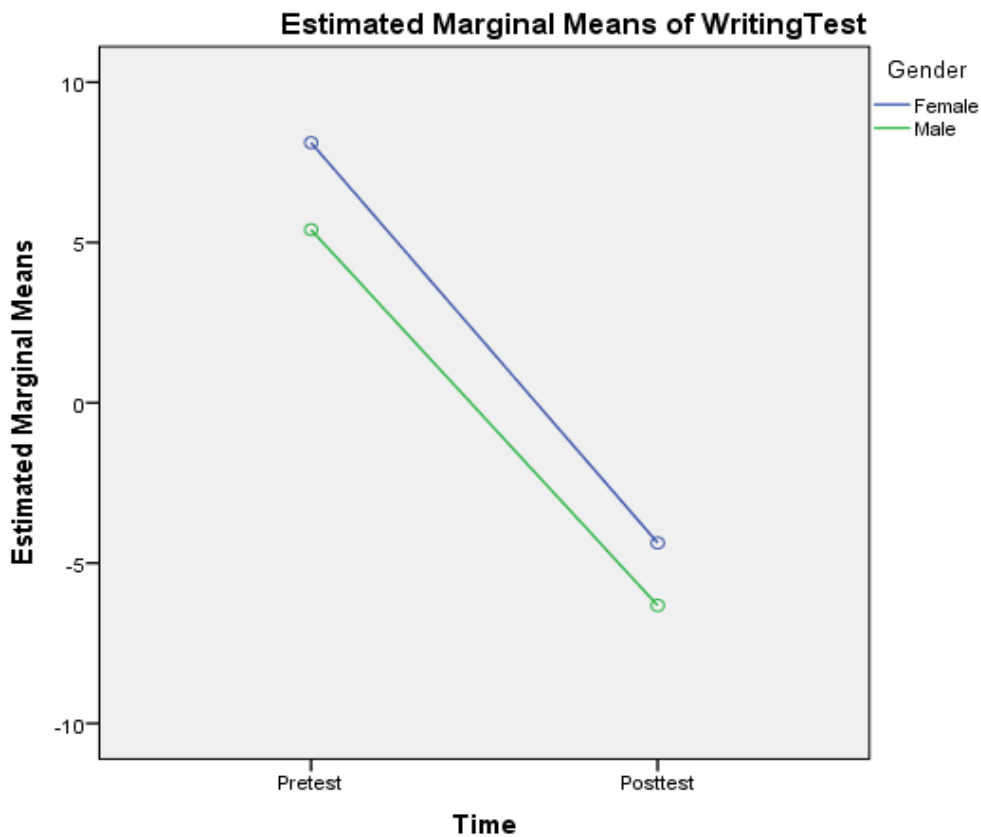


Figure 8.2 Graph of Writing Test by Gender

Table 8.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	8.11	6.72	349	5.40	5.68
Posttest	431	-4.37	3.65	349	-6.32	3.35

Figures 8.1 and 8.2, and Tables 8.1 and Table 8.2 indicate that girls have performed slightly better than boys, but not by much. This means that the difference was not significant.

Results for Two-Way ANCOVA

The Two-Way ANCOVA results for the Writing Test are given as follows. This was administered to assess the comparison of two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught), and gender (male and female students). The dependent variable was scores on the posttest, administered after the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language taking place (Time 2). Scores on the pretest administered prior to the start of the programs (Time 1) were used as a covariate to control for individual diversity.

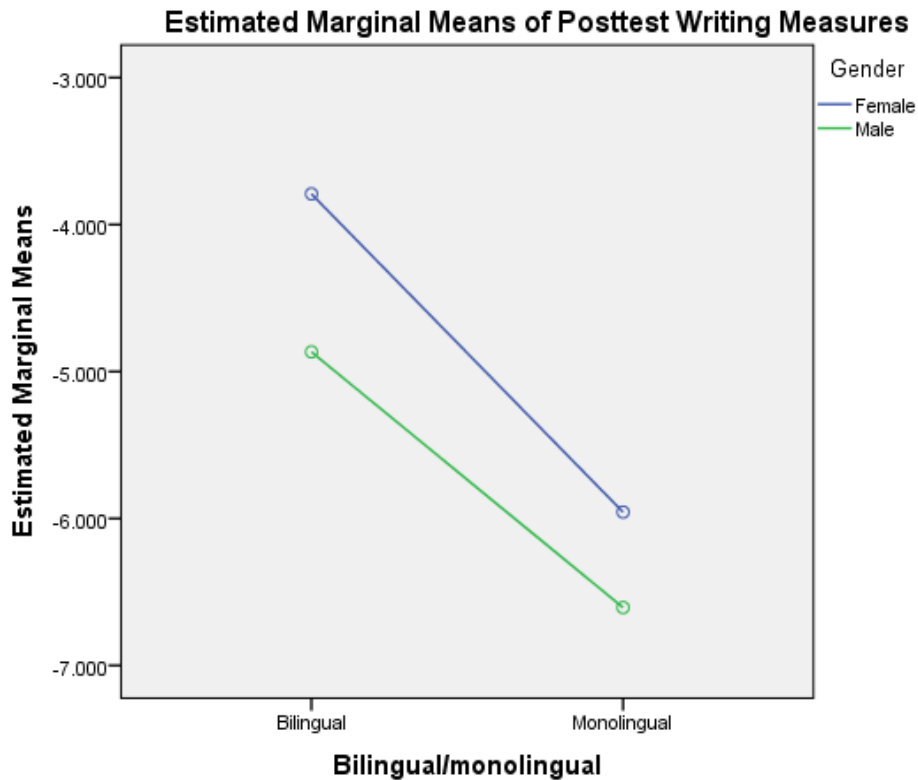
Table 8.3
Interaction Effect and Main Effect of Writing between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest Writing Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared (η_p^2)
Corrected Model	7728.990 ^a	4	1932.248	571.979	.000	.747
Intercept	15459.092	1	15459.092	4576.162	.000	.855
Pretest RC Rasch Measure on Posttest Scale						
Bilingual monolingual	421.216	1	421.216	124.687	.000	.139
Gender	135.485	1	135.485	40.106	.000	.049
Bilingual monolingual * Gender						
Gender	8.632	1	8.632	2.555	.110	.003
Error	2618.089	775	3.378			
Total	31791.159	780				
Corrected Total	10347.079	779				

a. R Squared = .747 (Adjusted R Squared = .746)

In order to confirm that there were no violations of the assumption of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable

measurement of the covariate, preliminary checks were carried out and found to be satisfactory. After adjusting for the pretest Writing measures, the result confirmed that there was no significant interaction effect between the type of programs (bilinguals and monolinguals): $F(1, 775) = 2.555, p > .005 (.110)$, with a small effect size (partial eta squared = 0.003) (see Table 8.3 for pretest-posttest measures, and Figure 8.3 for plot).

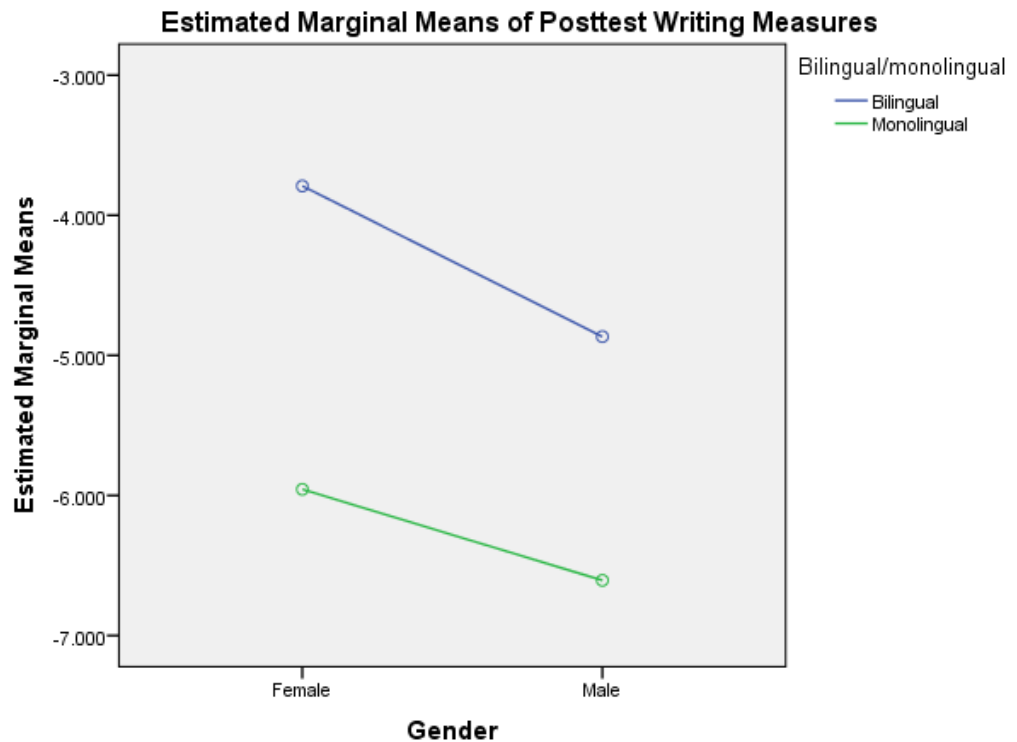


Covariates appearing in the model are evaluated at the following values: Pretest Writing Rasch Measure on the Posttest Scale = 6.90

Figure 8.3 Graph of Writing Test by Group (Bilingual/Monolingual) with Pretest as Covariate

On the other hand, the main effect for the language program was statistically significant with $F(1, 775) = 124.69, p = 0.000$, with $(\eta_p^2) = 0.139$ (considered a large effect size according Cohen’s 1988 regulations); as well as for gender, with $F(1, 775) = 40.106, p = 0.000$ with $(\eta_p^2) = 0.49$ (a very large effect size) (see Figure 8.4 for plots).

These results disclose that female students' score on the Writing Test was better than those of male students; and female and male students in the bilingual program responded better than female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest Writing Rasch Measure on the Posttest Scale = 6.90

Figure 8.4 Graph of Writing Test by Gender with Pretest as Covariate

Summary

The same English Writing pretest and posttest were given to bilinguals (N = 394) and monolinguals (N = 386) within a two month period (equivalent to 16 meetings), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA with the SPSS computer program and the pretests were used as the covariate. There are two main sets of findings: (1) with regards to the Mixed Design, and (2) with regards to the Two-Way ANCOVA.

The main findings, with regards to the Mixed Design ANOVA, are:

1. There was a significant interaction between the type of language program (bilingually-taught program and monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 0.818, $F(1,778) = 173.41$, $p=0.000$, partial eta squared = 0.182 (big effect).
2. There was a main effect for time (pretest and posttest), with Wilks' Lambda = 0.79, $F(1,778) = 9089.87$, $p = 0.000$, partial eta squared = 0.921 (very large effect).
3. The main effect comparing the two types of intervention was significant, with $F(1,778) = 720.70$, $p = 0.000$, partial eta squared = 0.481 (large effect).
4. There was no statistically significant interaction between gender and time, with Wilks' Lambda = 0.991, $F(1,778) = 7.378$, $p = 0.007$, partial eta squared = 0.009 (small effect).
5. The main effect for time (pretest and posttest) was significant, with Wilks' Lambda = 0.95, $F(1,778) = 7393.32$, $p = 0.000$, partial eta squared = 0.905 (very large effect).
6. The main effect comparing the two types of intervention was significant, with $F(1,778) = 47.614$, $p = 0.000$, partial eta squared = 0.58 (large effect).

The main findings, with regards to the Two-Way ANCOVA, are

1. There was no significant interaction effect between the type of programs (bilinguals and monolinguals): $F(1, 775) = 2.555$, $p > .005$ (.110), with partial eta squared = 0.003 (small effect).
2. The main effect for the language program was statistically significant, with $F(1,775) = 124.69$, $p = 0.000$, partial eta squared = 0.139 (small effect).
3. The main effect for the gender was significant, with $F(1,775) = 40.106$, $p=0,000$ with partial eta squared = 0.49 (large effect).

The results revealed that bilingual students were able to write English essay better than monolingual students. Even though there was no real distinction in students' ability on English Writing between females and males, female students from both

bilingual and monolingual schools scored better on the writing test than male students from bilingual and monolingual schools. The posttest results for English Writing were measured to be lower, after two months, than the pretest results, and this was true for both females and males, and for both bilinguals and monolinguals. This was probably because young Indonesian students find writing in English to be hard and the English Writing test was marked hard (at a high level). In other words, it requires longer than two months for significant improvements in English Writing to occur.

The next chapter explains the Rasch analysis with the RUMM2030 computer program for the Attitude/Behaviour Questionnaire.

CHAPTER NINE

DATA ANALYSIS (PART 5)

RASCH MEASUREMENT OF ATTITUDE/BEHAVIOUR

This chapter presents the findings of the Rasch analysis of the Behaviour posttest data.

The original Attitude/Behaviour Questionnaire (I= 42 items) was designed to collect data which, when analysed with the Extended Logistic Model of Rasch (Andrich, 1988a), could be used to create a linear measure of students' English Behaviour at the posttest stage of the experiment. The items were answered in three response categories; 'never or rarely' (scored 0), 'some of the time' (scored 1), and 'most or all the time' (scored 2). The data were analysed with the Rasch Unidimensional Measurement Model computer program (RUMM2030) (Andrich, et al., 2010) to create a unidimensional linear scale.

Initial Analysis

The analysis involved running the data through the RUMM2030 program and checking for non-fitting items and items with poor discrimination, which were then deleted, and re-running the program. After four analyses, all the 21 attitude items (except item 38) were deleted because of misfit to the measurement model. This was not consistent with the model used to develop the Behaviour Questionnaire that was based on many previous studies where attitude and behaviour were measured together (see Waugh, 2003, 2005, 2010a, 2010b). The RUMM program does not tell the researcher why an item doesn't fit the Rasch Measurement Model, just that it doesn't fit. It was difficult to see why the attitude items didn't fit the measurement model, but the students were in their first year of being bilingually-taught and, because they were not strong in English reading, the misfit may have been primarily due to their low command of reading and understanding of English, related to the different classroom culture in the two groups, in a test situation. Thus, it may be that the two groups of students (bilingually-taught and monolinguals) did not have agreement on the difficulties of the items because of their differences in their command of English, combined with some

difference in culture (living in an Indonesian culture, and learning English culture through English lessons), and that was the substantial cause of the misfit. The fifth analysis, consisting of ten behaviour items (items 2,8,10,12,16,18,26, 28, 36 and 38) and one attitude item (item 7), produced a good fit to the measurement model. Deletion of the attitude item 7 (“*I say new words several times in English*”) and a re-analysis with the ten behaviour items produced a worse fit to the measurement model and so the attitude item 7 was re-instated. The following material shows the output from the RUMM program when a good, unidimensional, linear scale of Behaviour with respect to Learning English was created with 11 items.

Output from Final Analysis

Standardised Fit Residuals

The Fit Residual data for both items and students have a mean near zero and a standard deviation near one when the data fit the measurement model. The mean for items is 0.195, and for persons is -0.306. The Standard Deviation is 1.112 for items and 1.525 for persons. It shows that the data fit the measurement model satisfactorily and this means that there is a good consistency of item-student response pattern. Table 9.1 shows item and person fit to the measurement model for the Behaviour measure.

Table 9.1
Overall Fit Statistics for the Behaviour Measure (N=779, I=11)

	ITEM-PERSON INTERACTION			
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	0.195	0.458	-0.306
Standard Deviation	0.217	1.112	0.972	1.525

Notes:

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model.

Dimensionality and Item-Trait Interaction

The item trait interaction chi-square was 103.82, df=99, and p=0.35 (see Table 9.2). This indicated that there was good agreement amongst the students about the item difficulties all along the scale. This, in turn, means that a single parameter for each

student, the person measure, and a single parameter for each item, the item difficulty, can be used to accurately predict each student's response to each item. A principal components analysis of the residuals showed that the first eigenvalue was 1.45, which is within the chance value, supporting the view that a unidimensional measure has been made.

Table 9.2
Item-Trait Interaction for Behaviour Scale

Total Item Chi-Square	103.82
Separation Index	0.71
Total Degree of Freedom	99.00
Total Chi-Square Probability	0.35
Cronbach Alpha	0.73

Notes:

1. The Index of Person Separation (Andrich & Van-Schoubroeck, 1989) is interpreted like a Cronbach Alpha (Cronbach, 1951) and is good.
2. The item-trait interaction test is a chi-square. The results indicate that there was a good collective agreement amongst the students about the item difficulties along the scale.
3. All numbers are given to two decimal points because the errors are only up to two decimal points.

Person Separation Index

For a good measure, a Person Separation Index is desired to be 0.75 or greater, as it is an indicator that the student measures are separated by more than their standard error. The Person Separation Index for this study is 0.71 (see Table 9.2 above). It indicates that there is a reasonable separation of measures in relation to the error (which is about 0.08 logits). While the Person Separation Index is calculated on the Rasch-created parameters, the Cronbach Alpha (Cronbach, 1951) is calculated on the raw scores. In this case, the Cronbach Alpha was 0.73, indicating reasonable reliability of the data.

Individual Item Fit

The RUMM 2030 program calculates individual item fits to the measurement model. For the Attitude and Behaviour measure, all 11 items fit the measurement model (see Table 9.3). In addition, all the standardized residuals fall within the range -1.2 to +1.1, supporting the good fit to the measurement model (which usually has to be within plus or minus 2 SDs).

Table 9.3*Locations, Standard Errors, Residuals and Chi-Squares for Behaviour Items*

Item No.	Location	SE	Residual	DF	Chi-Square	DF	Probability
2	-0.270	0.069	0.471	706.27	5.065	9	0.83
7	-0.173	0.056	0.946	706.27	8.522	9	0.48
8	0.133	0.060	-0.098	706.27	4.401	9	0.88
10	0.129	0.059	0.699	706.27	15.060	9	0.01
12	0.382	0.059	0.030	706.27	7.224	9	0.61
16	-0.281	0.059	-1.133	706.27	14.555	9	0.10
18	0.101	0.058	0.314	706.27	5.077	9	0.83
26	0.065	0.059	0.229	706.27	6.843	9	0.65
28	0.185	0.057	0.563	706.27	4.026	9	0.91
36	-0.481	0.059	-0.225	706.27	11.377	9	0.25
38	0.211	0.057	1.072	706.27	9.999	9	0.35

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.
3. DF means degrees of freedom. Probability is based on the chi-square fit to the measurement model.

Threshold Values

Items thresholds are positions on the scale between adjacent response categories where the odds are 1:1 that students will respond to a particular item, in either category. It is expected that the students would use the thresholds in the way that they were intended by the researcher and so the thresholds should be ordered in line with the conceptual ordering of the response and scoring categories. The thresholds in this measure were ordered in line with the conceptual ordering of the response and scoring categories (see Table 9.4).

Table 9.4
Item Thresholds Uncentralised (Item=11, Number=779) for Behaviour Measure

Item	Item Location	Thresholds	
		1	2
2	-.222	-1.805	1.361
7	-.132	-.804	.539
8	.157	-1.188	1.504
10	.124	-1.038	1.287
12	.352	-.623	1.327
16	-.142	-1.144	.858
18	.061	-.935	1.058
26	.092	-.975	1.160
28	.030	-.831	.893
36	-.439	-1.109	.230
38	.117	-.831	1.066

Note: The thresholds are ordered in line with the scoring categories.

Scoring Category Curve

The RUMM 2030 program produces curves of the scoring categories for each item. The Scoring Category Curves show the relationship between the probability of scoring in each category. Each item has three response categories: ‘Never or rarely’ (scored 0); ‘Some of the time’ (scored 1); and ‘Most or all the time’ (scored 2). The Scoring Category Curves should show a consistent relationship between the probability of scoring and the measure from low to high indicating that the scoring was done consistently and logically. A Scoring Category Curve for Item 2 is shown in Figure 9.1.

Figure 9.1 (Item 2: “*I pay attention to someone speaking English*”) showed that the scoring was done logically and consistently. When students have low measures on item 2, then they have a high probability of obtaining a zero score (the lowest response); and when they have a medium measure, they have a medium probability of scoring 1 (the moderate response); and when they have a high measure, they have a high probability of scoring 2 (the highest response). The Scoring Category Curves for the other items were checked and they showed logical and consistent scoring as well.

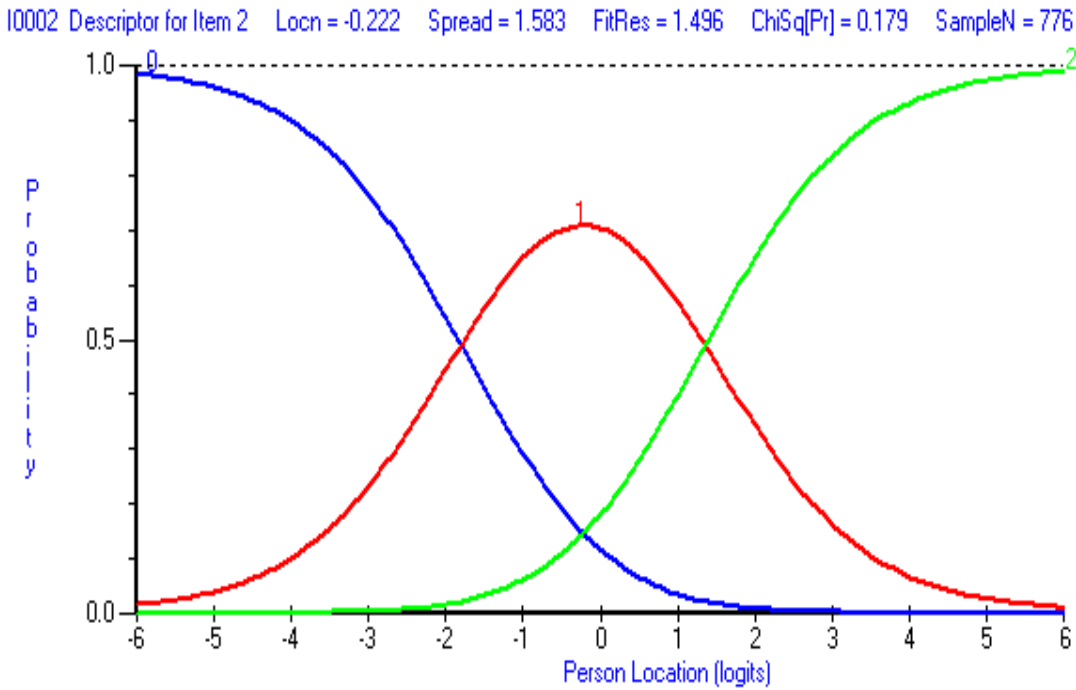


Figure 9.1 Item Category Curve for Item 2

Note: The blue curve is for a score of 0, the red curve for a score of 1 and the green curve for a score of 2.

Item Characteristic Curves

The Item Characteristic Curve provides information on item differentiation between persons and the item location. A group of students is considered to have performed well if their values (in the form of black dots) fit on the ogive curve. The ogive curve is the expected values for an item against the student measures (low to high). Take as an example Item 2 (Figure 9.2). With most of the dots close to the curve, it is considered a good fit to the measurement model and shows good discrimination. The Item Characteristic Curves for the other items were checked and found to be satisfactory.

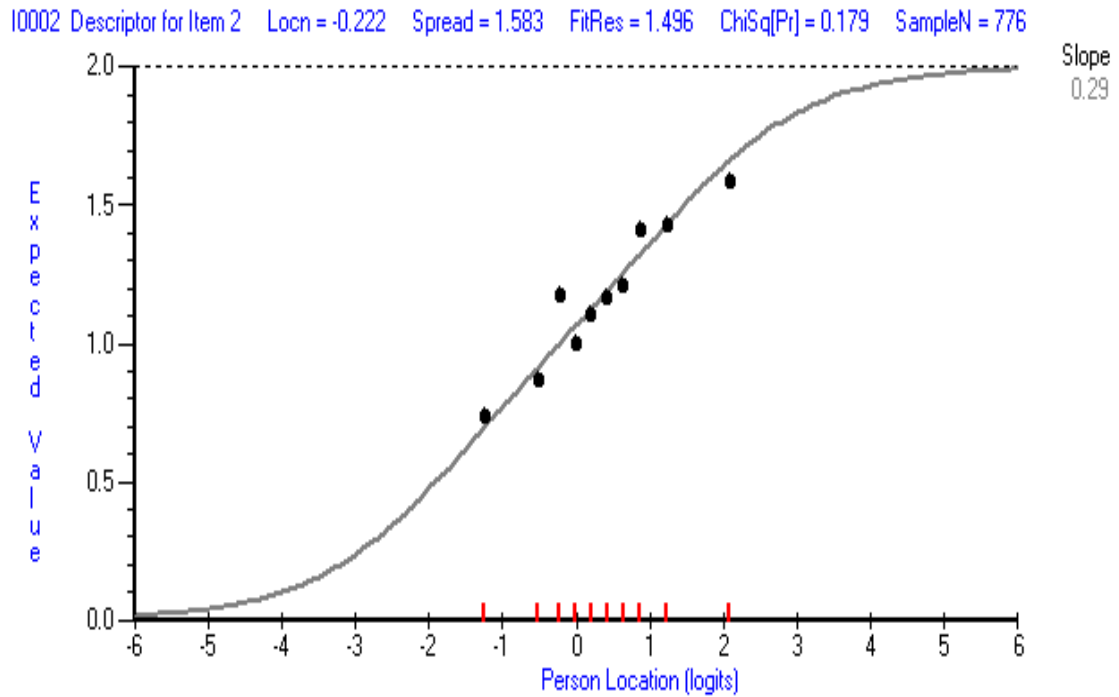


Figure 9.2 Item Characteristic Curve for Item 2

Person-Item Threshold Distribution (Targeting)

Measures are considered well-targeted when the range of item thresholds and the range of student measures are about the same. It means that the items are not too hard and not too easy for the students. Figure 9.3 shows the targeting graph for the measure of Behaviour. The targeting of the items is not as good as expected – there were insufficient easy, medium difficulty and hard items – but this is because many of the original items did not fit the measurement model and were deleted in the initial analysis. In the initial design of the items, there was a wide range of item difficulties but, apparently, the bilinguals and the monolinguals did not agree on the difficulties for most items, causing them to be deleted from this analysis.

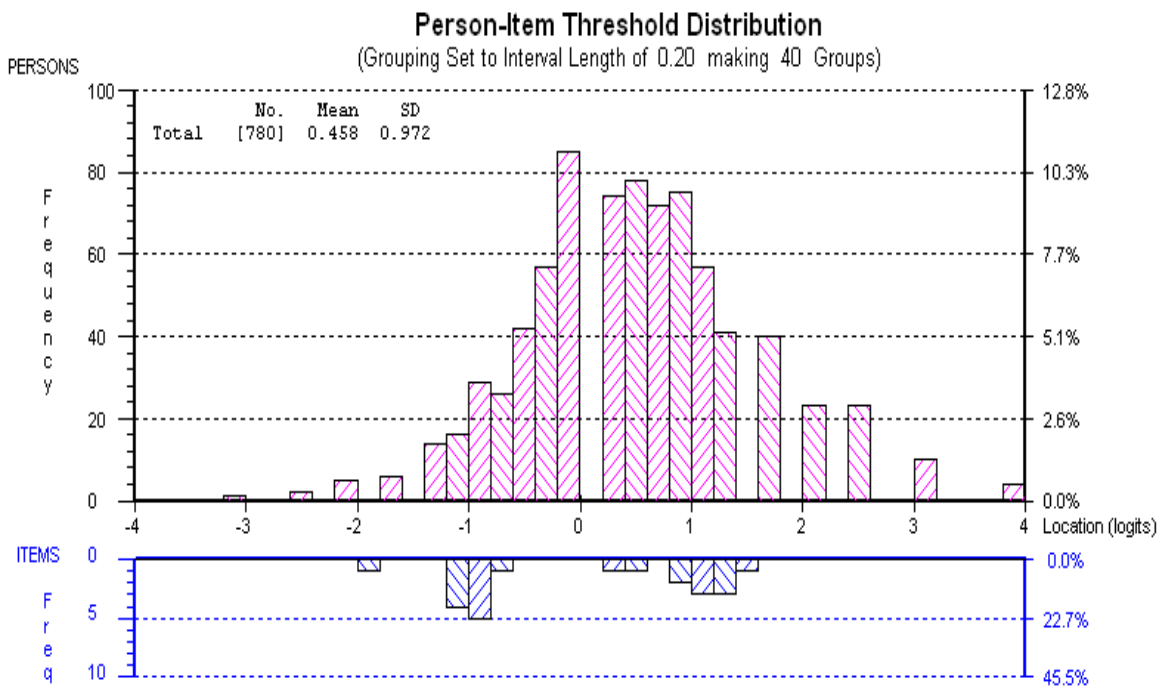


Figure 9.3 Person-Item Threshold Distribution for Behaviour

Figure 9.4 shows the distribution of item thresholds and the students on a ‘map’ and this also shows the restricted range of item thresholds. Somehow, this targeting problem would have to be rectified in any future use of the scale for these students.

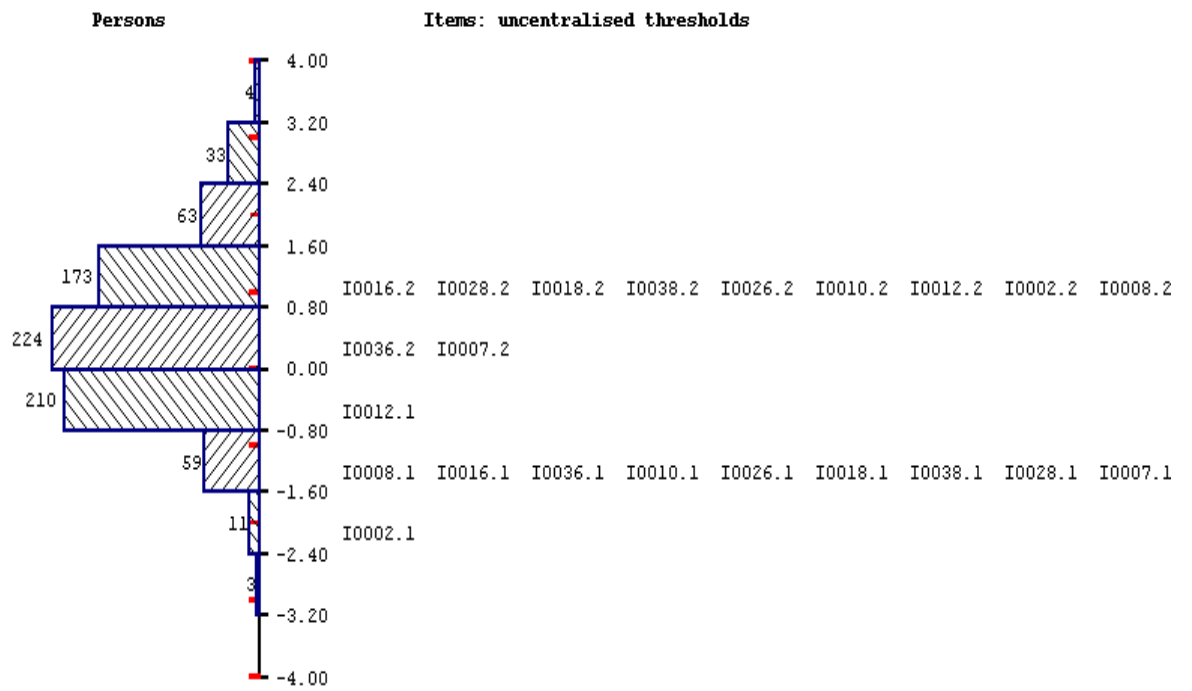


Figure 9.4 Behaviour Map

Note: I0016.2 means threshold 2 for item 16, I0002.1 means threshold 1 for item 2, and so on.

DIF by Gender

Each of the 11 items of the Behaviour measure showed no statistically significant Differential Item Functioning (DIF) by gender (see Figures 9.5, 9.6 and 9.7, for examples).

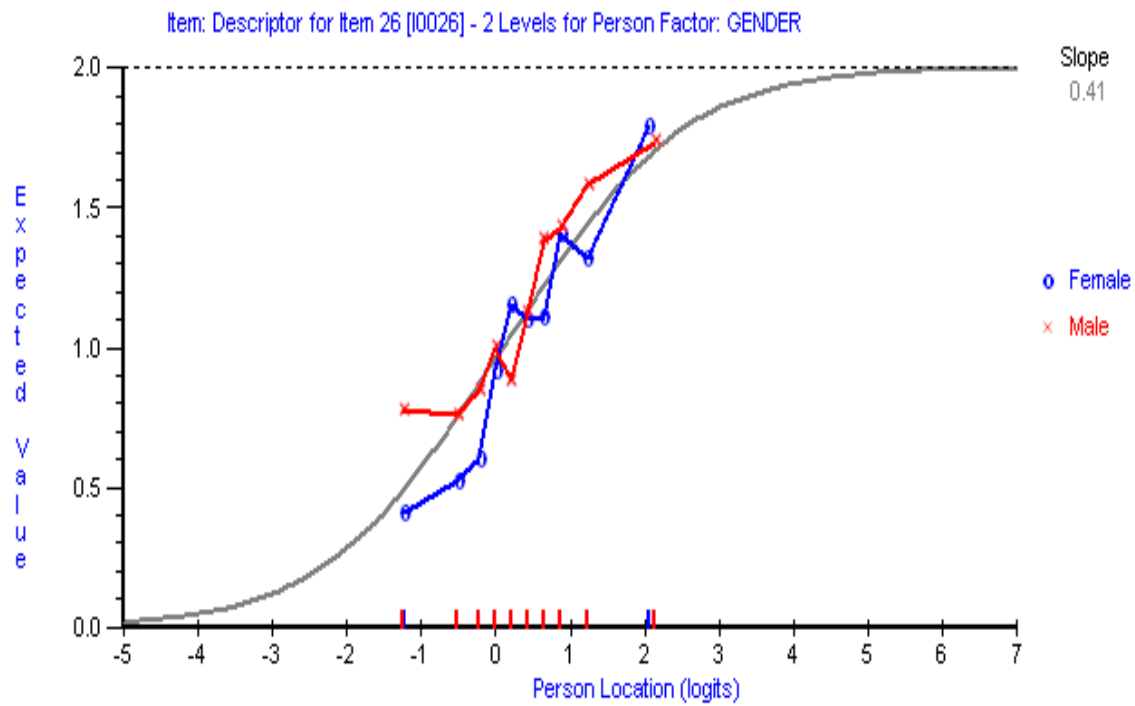


Figure 9.5 Item Characteristic Curves by Gender for Behaviour Item 26

Note: No statistically significant main effect by gender, $F=5.02$, $df=19,1$, $p=0.025$.

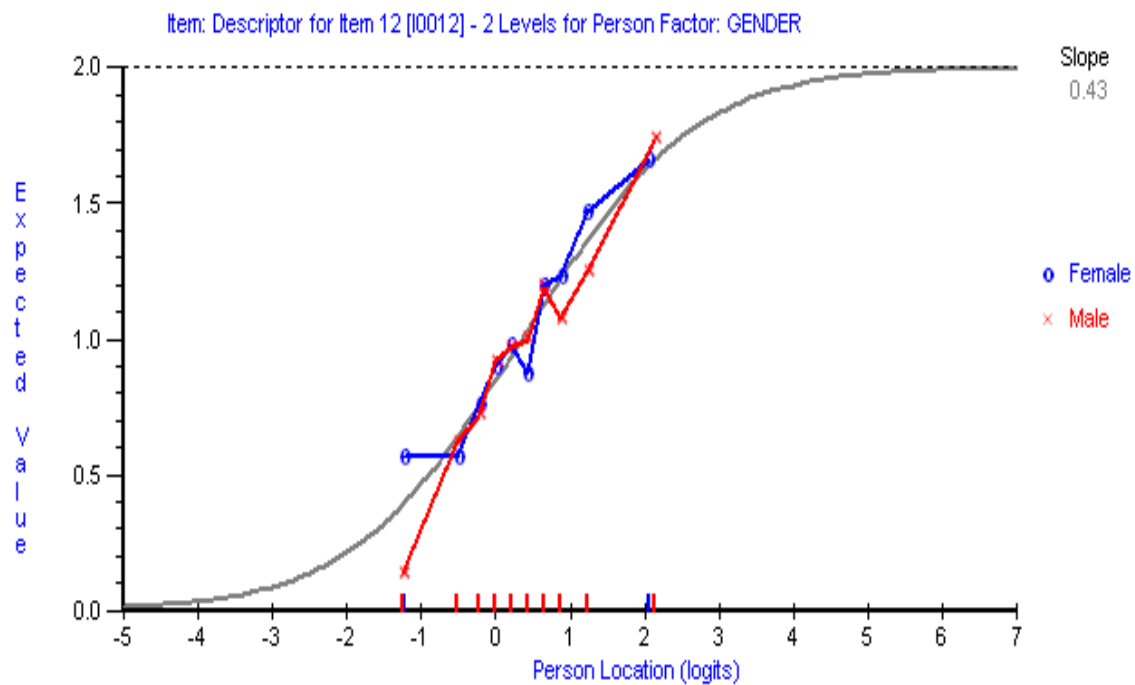


Figure 9.6 Item Characteristic Curves by Gender for Behaviour Item 12

Note: No Statistically significant main effect by gender, $F=2.96$, $df=19,1$, $p=0.09$.

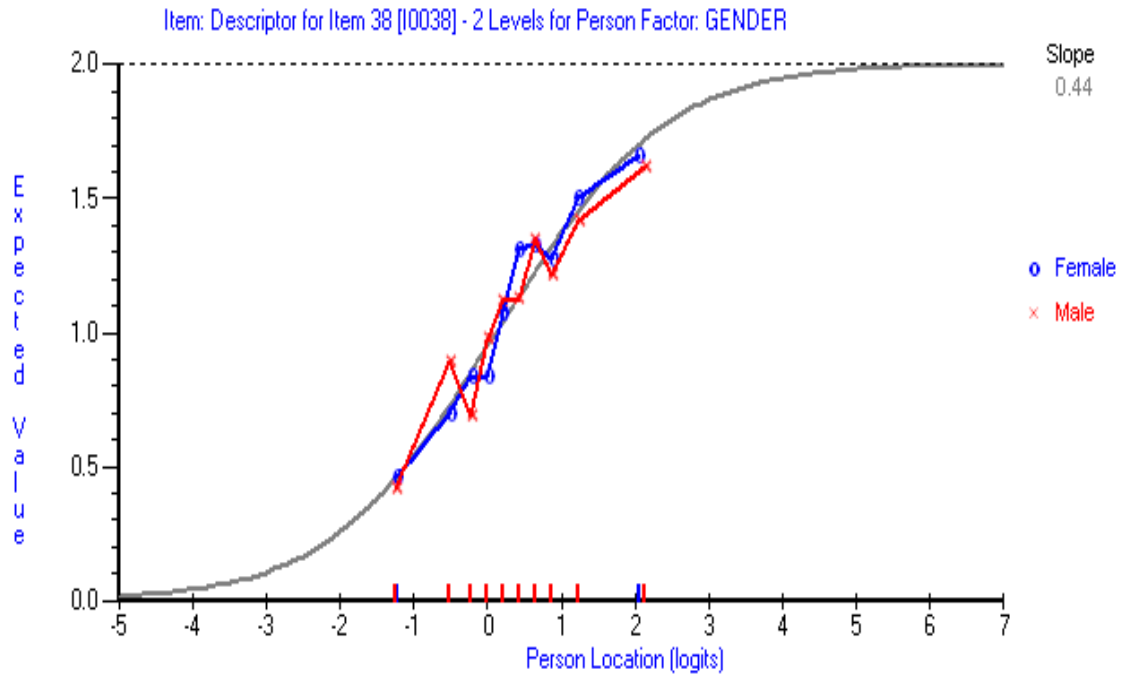


Figure 9.7 Item Characteristic Curves by Gender for Behaviour Item 38

Note: No statistically significant interaction effect by gender, $F=0.19$, $df=19,1$, $p=0.66$

All the Item Characteristic Curves for the other items were checked and showed no statistically significant difference by gender.

DIF by Type of Language Instruction

Only one item (Item 7) showed DIF by type where bilinguals had improved results on the Behaviour measure compared to monolinguals ($F=23.81$, $df=19,1$, $p=0.00000$, see Figure 9.8). Figures 9.9 and 9.10 show that there was no DIF for item 2 and item 38. However, over the Rasch measures for all the 11 items together, bilinguals had a statistically significantly higher Behaviour measure than monolinguals ($F=20.56$, $df=1,778$, $p=0.0000$, see Figure 9.12).

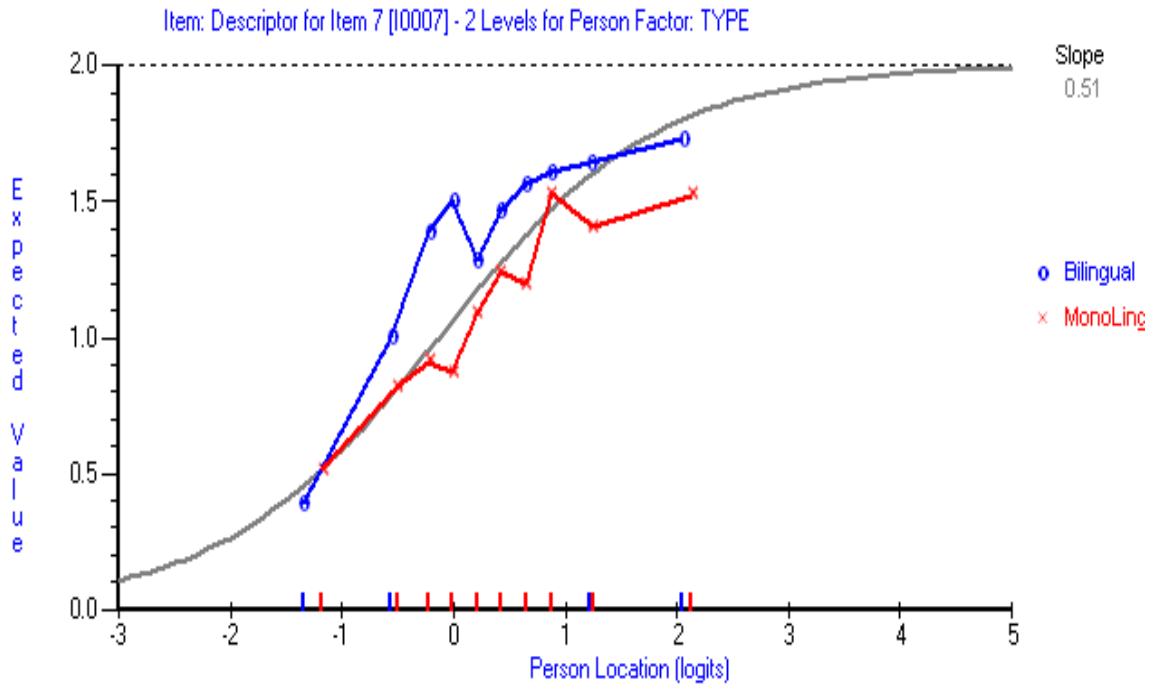


Figure 9.8 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 7

Note: There is a statistically significant main effect by type, $F=23.81$, $df=19,1$, $p=0.00000$

Bilinguals have a statistically significantly higher Behaviour measure.

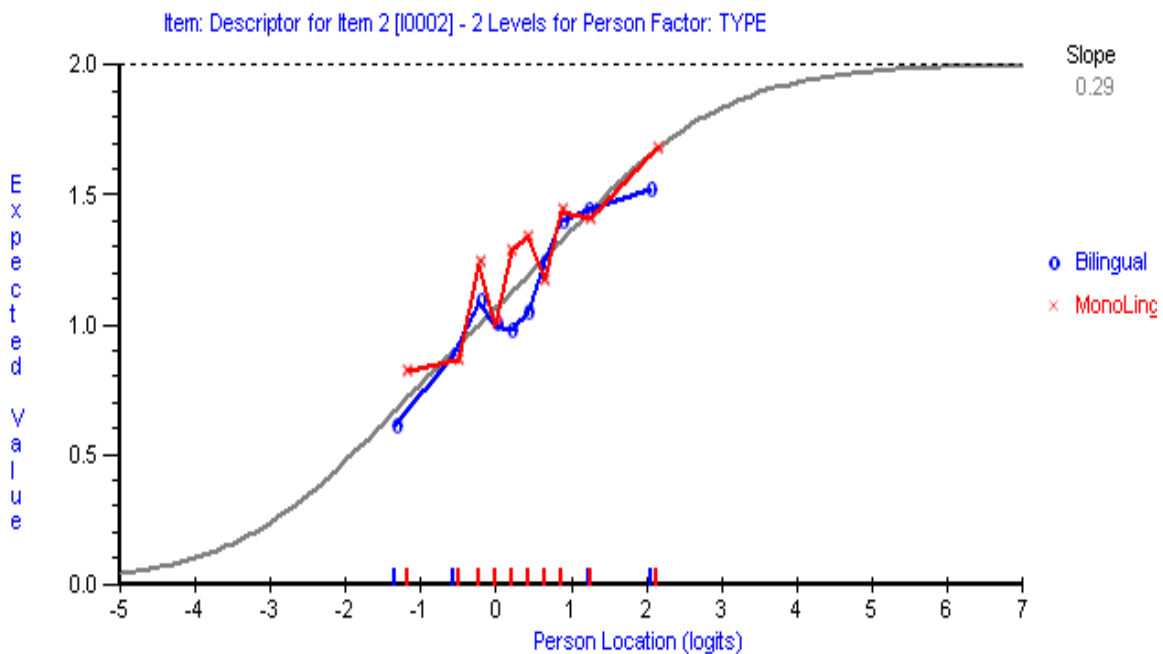


Figure 9.9 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 2

Note: Not statistically significant by type of teaching ($F=5.65$, $df=19,1$, $p=0.02$)

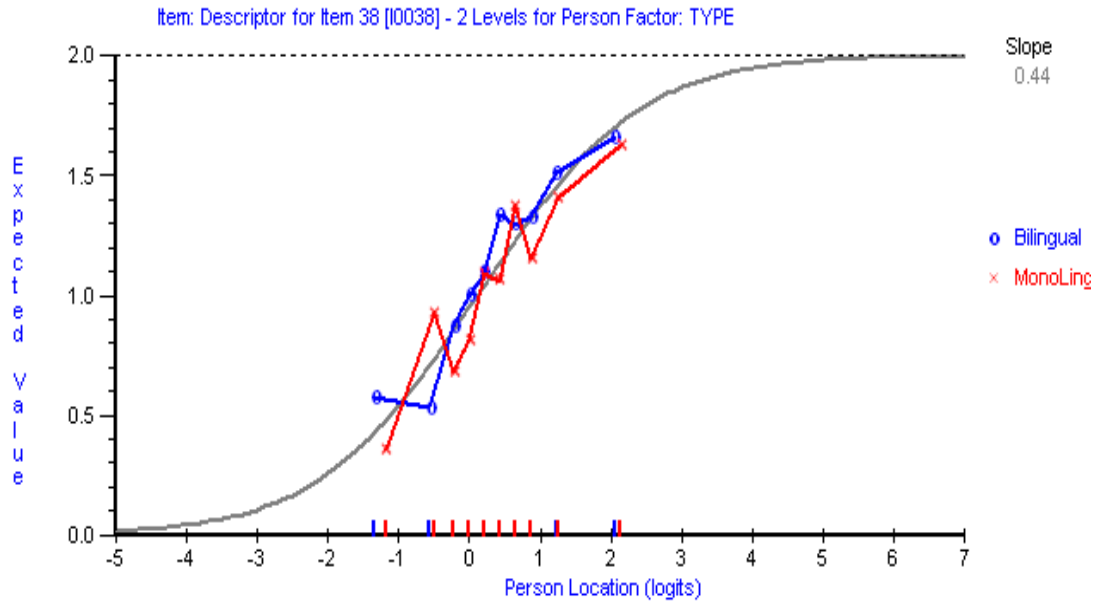


Figure 9.10 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 38

Note: Not statistically significant by type of teaching $F=4.66$, $df=19,1$, $p=0.03$

Targeting by Gender and Type of Language Instruction

Females showed a statistically significantly better Behaviour measure than males ($F= 11.52$, $df=1,778$, $p=0.0007$, see Figure 9.11) and bilinguals have a statistically significantly better Behaviour measure than monolinguals ($F= 20.56$, $df=1,778$, $p=0.00002$, see Figure 9.12).

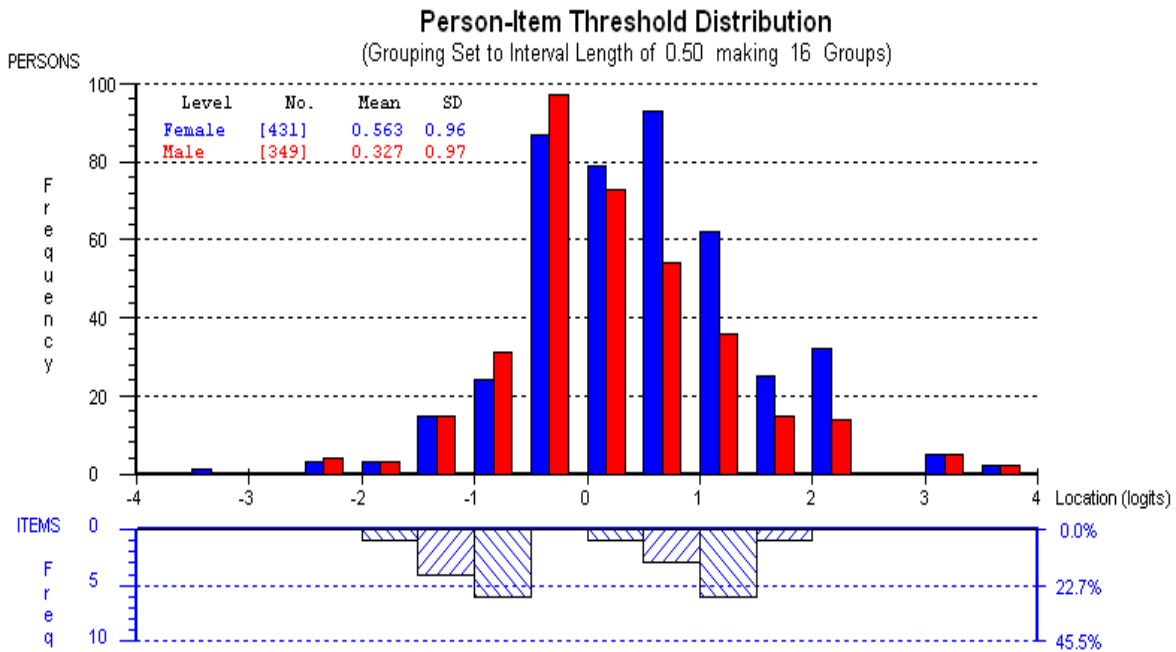


Figure 9.11 Targeting of Behaviour by Gender

Notes:

1. The person measures are on the upper-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the lower-side side from easy (LHS) to hard (RHS). $F= 11.52$, $df=1,778$, $p=0.0007$, which is a statistically significant difference.

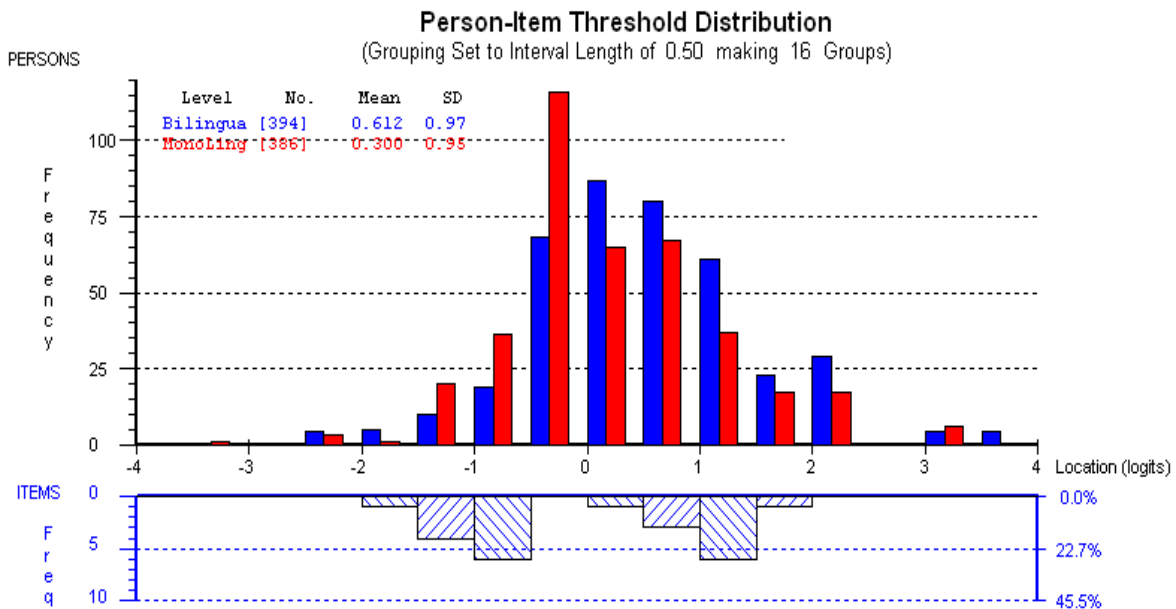


Figure 9.12 Targeting of Behaviour by Type of Language Instruction

Note: $F= 20.56$, $df=1,778$, $p=0.00002$ which is statistically significant for bilingual teaching.

Scale of Item Difficulties

Table 9.5 shows the item wording for Behaviour. The items have been ordered by difficulty from the easiest to the most difficult on the linear Rasch-created scale. The ordering of the items is consistent with the initial predicted conceptualized order, supporting the construct validity of the scale.

Table 9.5
Order of Difficulty of Items on the Linear Scale

Item No.	Item Location	Item Statements on the Behaviour
36 (easy)	-0.44	I actually like the way my teacher teaches English writing.
2	-0.22	I actually pay attention to someone speaking English.
16	-0.14	I actually read carefully words in English.
7	-0.31	I wish to say new English words several times.
28	0.01	I actually can understand English better when I do activities with friends.
18	0.06	I actually can read English at home on my own.
26	0.09	I actually learn more when I study English in groups.
38	0.12	I actually like English because we use it in the classroom.
10	0.13	I actually practice English with other students.
8	0.16	I actually say new English words several times.
12 (hard)	0.35	I actually start conversation in English with my friends.

Note: Item difficulties (locations) are measured in logits, the log odds of answering successfully.

The items were ordered from easy to hard on a linear scale (see Table 9.5) so that it can be seen which items are easy and which are hard. The easiest item involved the students' preference for the way their English teachers teach in the classroom, as expected (item 36 difficulty = -0.44 logits). The hardest item involved oral skill in ability to initiate speaking in English with their friends, as expected (item 12 difficulty = +0.35 logits).

Discussion and Summary

This chapter presents the results of the data analysis for the process of student behaviour in learning English as a second language in Aceh. Eleven items from the original 42 items produced a linear, unidimensional measure (31 items were deleted). The Fit Residual data showed that there was a good consistency for the item-person

response pattern. The Item-Trait Interaction (dimensionality) indicated that there was good agreement about the item difficulties along the scale. The Person Separation Index indicated that there was good separation of measures in comparison to errors. All items fitted the Rasch measurement model. The threshold values and the Scoring Category Curves showed that the scoring categories were used consistently and logically. The Item Characteristic Curves showed reasonable discrimination. All these data support the view that a linear, unidimensional measure of Behaviour was created so that valid inferences could be made.

However, 21 items, initially considered to be conceptually valid, had to be deleted because of misfit to the measurement model, apparently because the bilinguals and the monolinguals did not agree on the item difficulties. The deletion of these items caused a targeting problem, where the final scale had insufficient items across the whole difficulty range of the student measures. Thus this problem needs to be investigated further so that there are sufficient easy, medium and hard items to cover the full range of student measures in any future use of this scale.

The Item Characteristic Curves showed that the large majority of the items had no statistically significant differential item functioning (DIF) by gender and by type of language of instruction (bilingually-taught and monolingually-taught). However, over the Rasch measure for all 11 items together, bilinguals had statistically significantly better Behaviour measure than monolinguals and females had a statistically significantly better Behaviour measure than males.

The next chapter explains the experimental comparison results – pretest versus posttest measures by control and experimental groups for Reading Comprehension – based on the linear Rasch-created measures.

CHAPTER TEN

DATA ANALYSIS (PART 6)

EXPERIMENTAL RESULTS FOR BEHAVIOUR

In this chapter, the linear Rasch measures for Behaviour for the bilingually-taught group, known as the experimental group, were compared with similar measures for the monolingually-taught group, known as the control group, in a pretest/posttest, control/experimental group design. The measures on the pretest have been equated with those on the posttest, so that the Behaviour measures for the pretest control group, the pretest experimental group, the posttest control group and the posttest experimental group are all on the same scale and are already comparable. The Mixed Design ANOVA with the computer program SPSS (Statistical Package for the Social Sciences) (Pallant, 2007) was conducted to access the comparisons between two different interventions (bilingually-taught program and monolingually-taught program), and to access effects of gender on scores/responses from the Behaviour Questionnaire at both the pretest and posttest. The Two-Way ANCOVA with SPSS was run to check if the independent variable interacted to predict the dependent variables. It was used to test for main and interaction effects with SPSS to compared bilinguals with monolinguals and females with males.

Results for Mixed Design ANOVA (Split-Plot Design ANOVA)

The Mixed Design ANOVA was conducted to compare the measure of Behaviour Questionnaire for the two language programs (bilingual program and monolingual program) and of gender (female and male). The independent variables were the type of program (bilingual program, monolingual program) and gender. Measures for the dependent variable were based on the Behaviour Questionnaire posttest, administered following completion of the two months teaching English as a foreign language. The Behaviour Questionnaire pretest was administered prior to the teaching and its measures were equated with the posttest measures and used as a covariate to control for individual differences.

A mixed between-within subjects analysis of variance (Mixed Design ANOVA or Split-Plot ANOVA) was conducted to assess the difference between the two interventions (bilingually-taught program and monolingually-taught program); and the comparison of gender, on participants' Behaviour measures, across two time periods (pretest and posttest).

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, homogeneity and sphericity. There was no significant interaction between the type of language program (Bilingually-taught program and Monolingually-taught program) and time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.27$, $p = 0.607$, partial eta squared = 0.00. Similarly, there was no main effect for time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.51$, $p = 0.697$, $(\eta_p^2) = 0.00$ for the Behaviour measure. However, the main effect comparing the two types of intervention was significant, with $F(1,778) = 29.131$, $p < 0.0005$, $(\eta_p^2) = 0.36$. According to Cohen's effect size guidelines, this main effect is large. This suggested a real difference in students' pretest and posttest in which bilingual students performed better in Behaviour measures than monolingual students (see Figure 10.2). It is noted that bilingual students started with a better responses on Behaviour measure pretest than monolingual students.

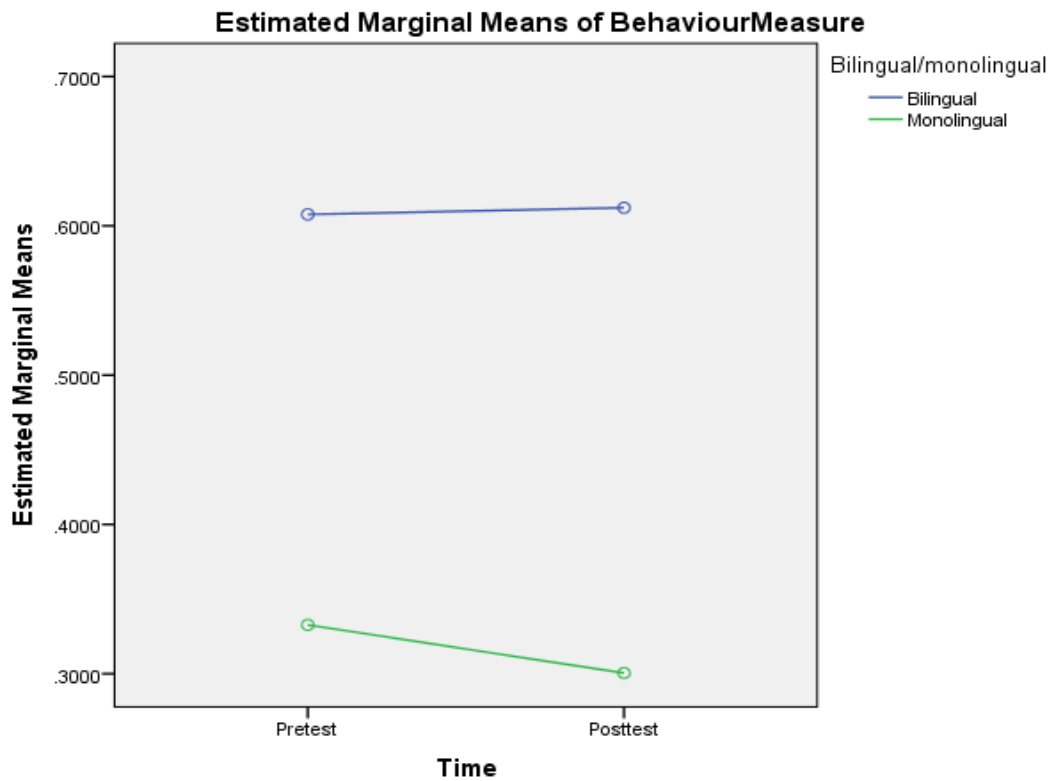


Figure 10.1 Graph of Behaviour by Group

For gender, there was no significant interaction between gender and time, with Wilks' Lambda = 0.996, $F(1,778) = 2.909$, $p = 0.89$, partial eta squared = 0.004 (a very small effect according Cohen's 1988 rules). Similarly, there was no main effect for time (pretest and posttest), with Wilks' Lambda = 1.0, $F(1,778) = 0.41$, $p = 0.839$, partial eta squared = 0.000 for the Behaviour measure. However, the main effect comparing the two types of intervention was significant, with $F(1,778) = 29.580$, $p < 0.0005$, partial eta squared = 0.37 (a large effect size according Cohen's 1988 regulations). This suggested a real difference in students' Behaviour measure, favouring females (see Figure 10.2 for plot).

Table 10.1
Pretest and Posttest Scores for the Bilinguals and Monolinguals

Time period	Bilinguals			Monolinguals		
	n	M	SD	n	M	SD
Pretest	394	.607	.874	386	.332	.831
Posttest	394	.612	.970	386	.300	.948

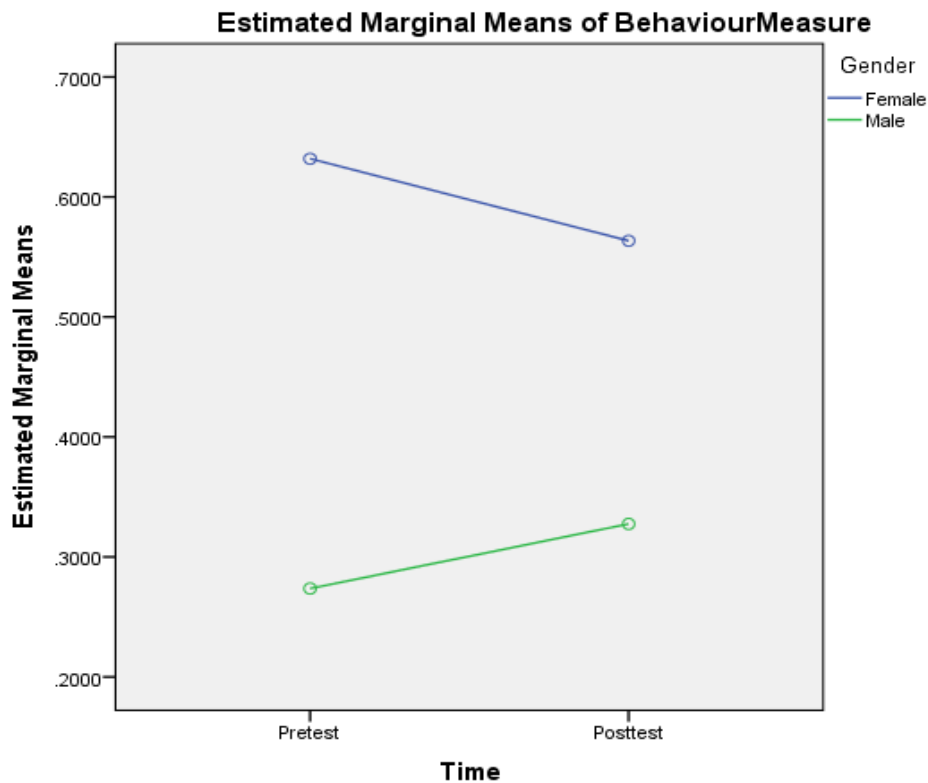


Figure 10.2 Graph of Behaviour by Gender

Table 10.2
Pretest and Posttest Scores for the Female and Male Students

Time period	Female students			Male students		
	n	M	SD	n	M	SD
Pretest	431	.63	.85	349	.27	.84
Posttest	431	.56	.96	349	.33	.97

Figures 10.1 and 10.2, and Table 10.1 and Table 10.2 suggest that females have performed better than males. It is well-known that the English achievement of girls is better than that of boys (for example, see Machin & McNally, 2005), but are girls' behaviour with regards to learning English better than boys, as well? The findings from this study suggest that it is. It may be the better behaviour of girls with regard to learning English that is important and that, at least partially, explains the well-known finding that the English achievement of girls is better than that of boys.

Results for Two-Way ANCOVA

The following are the Two-Way ANCOVA results for the Behaviour measures. This was conducted to compare the two programs: the bilingually-taught program and the monolingually-taught program for male and female students. The independent variables were the type of program (bilingually-taught program and monolingually-taught) and gender (male and female). The dependent variable was scores on the posttest, administered following the two months (equivalent to 16 meetings) of the teaching and learning of English as a foreign language (Time 2). Scores on the pretest administered prior to the commencement of the programs (Time 1) were used as a covariate to control for individual differences.

Preliminary checks were conducted to ensure that there were no violations of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for the pretest of the Behaviour measure, there was no significant interaction effect: $F(1, 775) = .104, p > 0.05 (.747)$, with a small effect size ($\eta_p^2 = 0.000$) (see Table 10. 3).

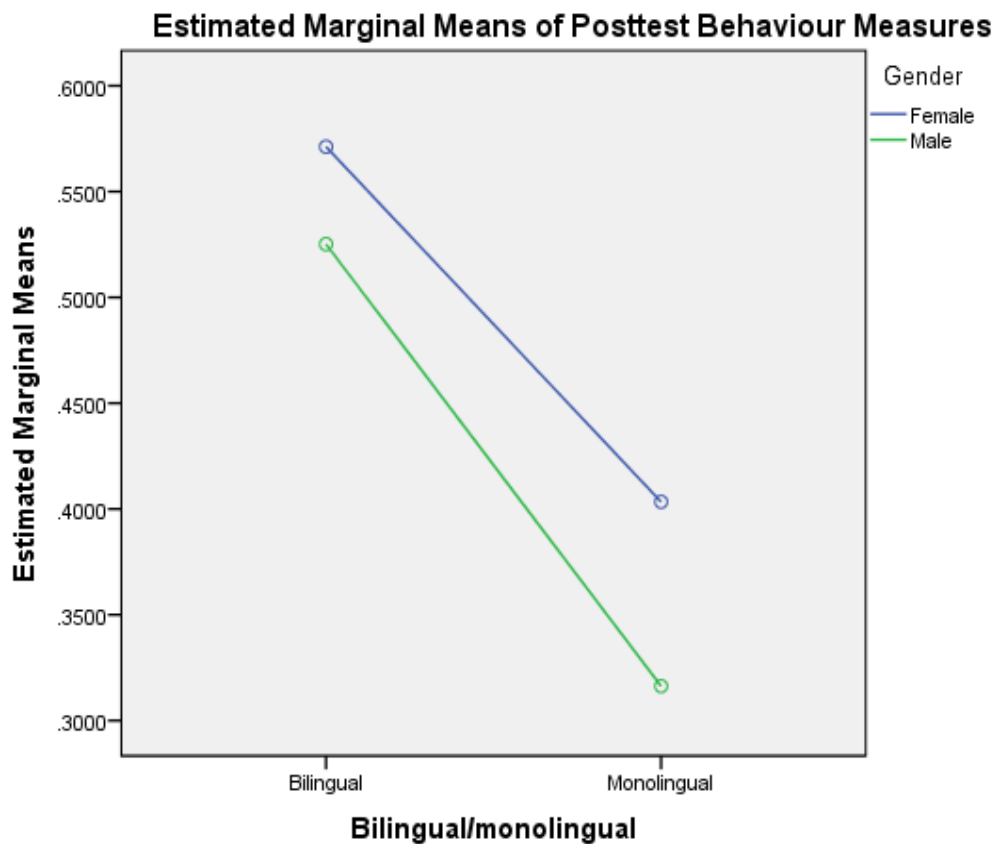
Table 10.3

Interaction Effect and Main Effect of Behaviour between the Bilingually-taught Group and the Monolingually-taught Group, and Gender

Tests of Between-Subjects Effects						
Dependent Variable: Posttest Behaviour Measures						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	134.304 ^a	4	33.576	43.280	.000	.183
Intercept	35.963	1	35.963	46.356	.000	.056
PretestBehaviourRaschMeasureon PosttestScale	106.191	1	106.191	136.881	.000	.150
Bilingualmonolingual	6.670	1	6.670	8.598	.003	.011
Gender	.818	1	.818	1.054	.305	.001
Bilingualmonolingual * Gender	.081	1	.081	.104	.747	.000
Error	601.235	775	.776			
Total	899.066	780				
Corrected Total	735.539	779				

a. R Squared = .183 (Adjusted R Squared = .178)

The main effect for the language program was statistically significant with $F(1,775) = 8,598, p = 0.003$; while that for gender was not, with $F(1,775) = 1,054, p = 0.305$ (see Figures 10.3 and Figure 10.4).

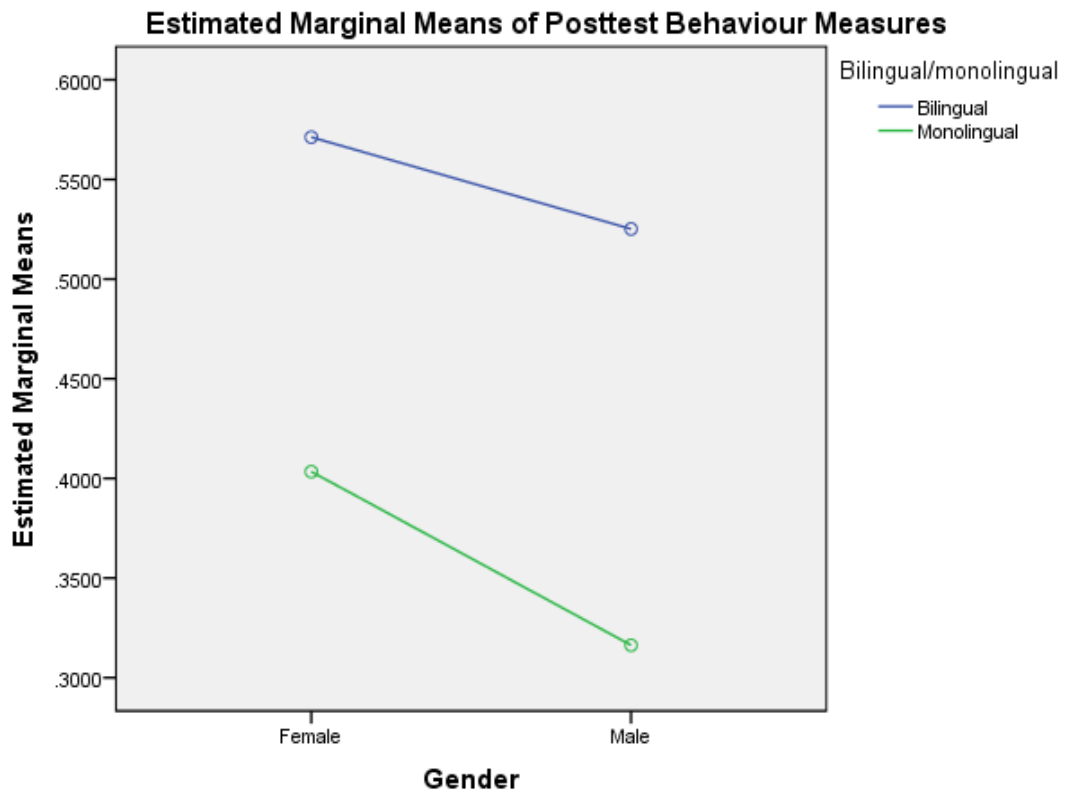


Covariates appearing in the model are evaluated at the following values: Pretest Behaviour Rasch Measure on Posttest Scale = .471579

Figure 10.3 Graph of Behaviour by Group (Bilingual/monolingual) with Pretest as Covariate

These ANCOVA results are different from the Split-Plot ANOVA results because the common covariate pretest measures were used to control for individual differences in monolingual and bilingual measures in one case, and in gender differences in the second case.

These results suggest that female students responded better than male students; and female and male students in the bilingual program responded better than female and male students in the monolingual program.



Covariates appearing in the model are evaluated at the following values: Pretest Behaviour Rasch Measure on Posttest Scale = .471579

Figure 10.4 Graph of Behaviour by Gender with Pretest as Covariate

Summary

Bilinguals (N = 394) and monolinguals (N = 386) students were given an English Behaviour Questionnaire pretest and posttest within a two month period (equivalent to 16 meetings), as part of learning English as a foreign language. Rasch measures were used in a Mixed Design ANOVA and Two-Way ANCOVA using the SPSS computer program. There are two main sets of findings: (1) with regards to the Mixed Design, and (2) with regards to the Two-Way ANCOVA.

The main findings for Behaviour with regards to Learning English, in respect of the Mixed Design ANOVA, are:

1. There is no significant interaction effect between bilinguals and monolinguals with $F(1, 775) = .104, p > .005 (.747)$, and a small effect size (partial eta squared = 0.000).
2. Bilinguals responded better than monolinguals.
3. Females responded better than males.
4. The present study suggested that the behaviour of girls with regard to learning English was important and that it, at least partially, may explain the well-known finding that English achievement of girls is better than that of boys.

The main findings for Behaviour with regards to Learning English, in respect of the Two-Way ANCOVA, are:

1. There was a main effect between bilinguals and monolinguals, with $F(1, 775) = 8,598, p = 0.003$, and a small effect size (η_p^2) = 0.011 (favouring bilinguals).
2. The main effect for gender was not statistically significant with $F(1, 775) = 1,054, p = 0.305$, and a small effect size (η_p^2) = 0.001.

The findings showed that bilingual students scored higher on the Behaviour measure than monolingual students. Female students from bilingual and monolingual schools scored higher on the Behaviour measure than male students from bilingual and monolingual schools.

The next chapter analyses the qualitative data on the students' written comments.

CHAPTER ELEVEN

DATA ANALYSIS (PART 7)

STUDENTS' WRITTEN COMMENTS

This chapter presents a qualitative analysis of the written comments on students' experiences with regards to learning English as a foreign language, using the Miles and Huberman (1994) method.

These data enrich the previous findings reported in Chapters Five, Six, Seven, Eight, Nine, and Ten. The present chapter analyses the opinions and thoughts of the students regarding their experiences in learning English prior to their participation in this study. Ten themes (or main issues) were identified: (1) Tasks on Learning English as a foreign language, (2) Student-Student Relationships, (3) Student-Teacher Relationships, (4) Personal Views on the Benefits of Learning English, (5) Common Views on the Benefits of Learning English, (6) Confidence and Achievement in Learning English, (7) Learning English through Media, (8) Family Support in Learning English, (9) Obstacles in Learning English, and (10) Other Views. These data give insights into how students feel with regard to their learning of English as a foreign language. Did they like learning English? What are the benefits of learning English? Did family, peers and teachers help them in learning English? Were there any obstacles in learning English? Did they have any support in learning English?

The questionnaire data that were collected between January and April 2011 included an additional written question as part of the posttest questionnaire. These written comments were provided in response to one written question on their questionnaire posttest, "***Is there anything else you would like to add about your experiences in learning English?***" Of the 780 students (bilinguals, N=394, bilinguals, N=386) who completed the posttest, 702 students (90%) answered this additional written question.

From the 702 students who answered the additional written question, 1842 written comments were provided: 1110 comments were from bilinguals and 732 were

from monolinguals. The comments, originally written in Bahasa Indonesia, were translated into English, back-translated to ensure the accuracy, and carefully grouped into ten areas and tally-marked manually. Tally marks are strokes written to record frequency of data, e. g. (////) and recorded as number of occurrences (see Appendix Q for the complete number of occurrences of written comments). A number of the students' best quotations that show comprehensive or unique information are provided here. By comprehensive information, it is meant that the comments were provided with details; and by unique, it is meant that the comments portrayed different senses and nuances that were not commonly depicted by other students. The students' comments were coded using their student numbers, which were allocated anonymously, and they are quoted verbatim.

It is worth noting that there are no gender-specific pronouns in Bahasa Indonesia. Both 'she' and 'he' in English are both referred to as '*dia*' in Bahasa Indonesia. On the written quotations, some students mentioned '*dia*'. Whenever the gender was unknown to the researcher, based on what the students provided, the pronoun 'she' or 'he' from the comments was translated to both 'she/he', for possessive pronouns to 'her/his'; and to objective pronouns to 'her/him'.

The analysis of the data discovered that both bilingually and Monolinguals found learning English as a foreign language is fun, yet challenging. However, they were motivated to learn it in order to help them for their future studies, for their future jobs and even to help them to go abroad and to study abroad. They thought that it is important to learn it, so they tried ways to make it easier to learn, despite obstacles along the way. Teachers helped in their journey of learning English.

Theme 1: Tasks on Learning English as a foreign language

There were four tasks on learning English as a foreign language identified: (1) Tasks for Listening, (2) Tasks for Speaking, (3) Tasks for Reading, and (4) Tasks for Writing.

Tasks for Listening

Some of comments from both groups stated that they liked listening to English songs (bilinguals, N=25; monolinguals, N=19). Other students said that they asked others to

speak slowly or repeat words in English (bilinguals, N=3); paid attention to someone speaking English (bilinguals, N=1; monolinguals, N=1); and that listening to English helped her/him to write (bilinguals, N=1).

Tasks for Speaking

More bilinguals than monolinguals (bilinguals, N=21; monolinguals, N=10) liked speaking in English. Eighteen comments showed that the students liked to practise speaking with other students, sang songs and believed that it was important to converse in English fluently.

Tasks for Reading

Some comments indicated that the students like learning English through reading (bilinguals, N=16; monolinguals, N=8). Only a small number of bilinguals, stated that they guessed the meaning of English words in the texts, and that they liked reading books/e-books in English.

Tasks for Writing

There were fewer comments on English Writing and only ten comments to cover five statements, affirming that students liked writing, wrote feelings in a diary, liked to create a new sentence with a new vocabulary, used a dictionary for new words, and understood written text better than spoken utterances.

The followings are students' quotations relating to Tasks on Learning English as a Foreign Language.

Students' quotations:

In our VII-1 class (seven-one), we were applying an English Day program which was held on Wednesdays. Sometimes, we forgot to speak in English, but we reminded each other. Our class was therefore called COSEVEN (Community of Seven-One). (Student # 1353, bilingual)

We have even got an English teacher from an English speaking country. The teacher taught me by using English language. I felt that I could understand him a little bit. From that time on I am motivated in learning English, so I will be able to communicate to people from other countries in English when I am overseas. (Student # 1022, monolingual)

So much so I wanted to be able to learn English, I sacrifice my afternoon time to take an after-school English tuition. And I do not mind to memorise English words which I do not know the meanings of. I also like to listen to songs in English. But I dislike getting questions in English, since I do not know their meanings. (Student # 1063, monolingual)

When I was having some walks, I would look at some items around me. I would try to name them in English. Should I do not know their names in English, I would seek help from my teacher or my friends. (Student #1384, monolingual)

I have been in a boarding school. Every day after *Shubuh* [an hour before sunrise] prayer, we were taught numbers of English words. I, then, used those words to communicate with my friends. I loved speaking English with my friends. I like it (Student # 1386, monolingual).

I was so happy when I communicated with some Dutch men in English (Student # 1109, bilingual).

Theme 2: Student-Student Relationships

Twenty-nine comments revealed that bilinguals learnt more when studying English in groups, and they asked friends to help them in learning.

Students' quotations:

I like English because I like to listen to it and learn it. Sometimes my friends and I read English books. We understand English and are able to write it and sing English songs. (Student # 1033, monolingual).

My experience in learning English is that I really want to speak and write well in English. But that is not happening yet. When I was in my primary school, my English lessons were not quite deep yet, so I asked a friend of mine to help me in English, because her/his English was quite good due to her/his school at one of favourite schools. (Student # 1021, bilingual).

I sometimes find English both confusing and fun. I know that it is impossible to understand it quickly, but it needs a process to learn it. I often seek help from my friends. (Student # 1048, bilingual).

I played games with my teacher and my friends. (Student # 1021, bilingual).

I have got a nice experience in learning English: I have performed an English dialogue with my friend in front of the class. (Student # 1116, bilingual).

I quite like English. I like it when I performed English dialogue with my friends in front of the class, and, I disliked it if I was to memorise in English. (Student # 1142, bilingual).

I have spoken in English with my friend. (Student # 1185, bilingual).

For me, learning English is lots of fun because my friends and I usually practice speaking in English. But sometimes, it turns out boring as well. (Student # 1188, bilingual).

My friends and I can speak English because we often practise speaking in English. My hope to study English is that in order it helps me to go overseas and study there. I also want to make my parents proud and happy. (Student # 1196, bilingual).

I got confused the first time I learn English. After some months I start to like it. And now I practise English quite regularly with my friends from the same year or from above. (Student # 1205, bilingual).

To sum up, both bilinguals and monolinguals who commented on this theme stated that they learnt English more when they studied it in groups. In addition to that, most of students commented regarding this theme asked friends to help them in learning English.

Theme 3: Student-Teacher Relationships

Many students from both groups (N=116) stated that they liked the way their teacher taught them English in that they learnt a great deal from their English teachers and they would seek help if they encountered problems. However, a small number (N=16) from both sides found the way their teachers taught English was not favourable to them. They also believed that the way teachers taught influenced their love or dislike of English.

Students' quotations:

Like the way English is taught

In my opinion, the way my teacher teaches me influences the way I learn English. Should the teacher be good, kind, and does not like to easily get grumpy, I like her. Frankly to say, should the teacher be mean and unkind; anything she/he has taught me will not reach my mind. (Student # 1134, bilingual).

My experience in learning English is enjoyable, because I like the way my English teacher teaches English. She makes me understand. (Student # 1364, monolingual).

I like my English teacher best. She is kind and not easy to get angry. (Student #1371, monolingual).

Dislike the way English is taught

I am a little bit confused with my English teacher because he does not understand English well. The way he writes, reads and explains English makes me dislike to be taught by him. When we were learning English, the teacher explains it in local language. We got mixed up. Therefore, I learn again at home or at my English tuition centre so my English knowledge and skills keep increasing. (Student # 1026, monolingual).

I like English but I do not like the way my teacher teaches it. Because of his teaching at my school, I do not like English anymore. (Student # 1028, monolingual).

As a matter of fact, I prefer overseas teacher to teach us English. It seems easier for me to understand, and it would be easier if the teacher can also speak in Bahasa Indonesia. (Student # 1243, monolingual).

Yes. Sometimes my teacher does not answer some questions that I asked. (Student #1363, monolingual).

As a summary, most bilinguals and monolinguals who commented on this theme liked the way their teachers taught English and only some did not like their teachers' methods. Some of them learnt a great deal from their English teachers and were pleased to state this but some other students believed that the way a teacher teaches influences whether students like or dislike learning English.

Theme 4: Personal Views on the Benefits of Learning English

Students commented that they liked learning English (bilinguals, N=96, monolinguals, N=70); that English was fun (bilinguals, N=76, monolinguals, N=50); that it was challenging (bilinguals, N=19, monolinguals, N=9); that learning English is important (bilinguals, N=9, monolinguals, N=12); that it was easy (bilinguals, N=15, monolinguals, N=5); that English was a unique language (which sounds nice but is is

difficult to pronounce) (bilinguals, N=7, monolinguals, N=9); that they were aware that a number of good texts were written in English (bilinguals, N=1); that they preferred to learn English with simple sentences (bilinguals, N=1, monolinguals, N=2); that English was their favourite subject (bilinguals, N=1); that learning English needed time (bilinguals, N=2, monolingual, N=1); that other languages helped them understand English (bilingual, N=1, monolingual, N=1); that they liked storytelling (bilinguals, N=2); that they liked creating stories in English (bilingual, N=1); that they liked to translate English into Bahasa Indonesia and vice versa (monolinguals, N=3); and that an English dictionary helped them (bilinguals, N=7, monolinguals, N=3) (see Table 11.1).

Table 11.1
Students' Personal Views on the Benefits of Learning English as a Foreign Language

Personal Views on Benefit of Learning English as a Foreign Language	Bilingual	Monolingual	Total
I like learning English	96	70	166
I like English because it is enjoyable/fun	74	50	124
I like English because it is challenging	19	9	28
Learning English is important	9	12	21
Learning English is easy	15	5	20
I like English because it is unique (it sounds nice but is difficult to pronounce)	7	9	16
Using a dictionary helps me	7	3	10
I am happy with my vocabulary	1	3	4
I want to learn English with simple sentences	1	2	3
Learning English needs time	2	1	3
I like to translate	0	3	3
English is enjoyable if studied correctly	2	0	2
Other languages help me understand English	1	1	2
I like storytelling	2	0	2
I like English because a number of good books are in English	1	0	1
English is my favourite subject	1	0	1
I like creating stories in English	1	0	1

Source: Prepared by the author from the students' written answers

Students' quotations:

I used to dislike English lessons. In every English session at my primary school, my English teacher was always angry with me. Now, I am at this Middle School with a bilingual program, so I start loving English. (Student #1274, bilingual)

First, I like English if it is used in daily life. Second, I like English if there is somebody who can translate it into Bahasa Indonesia well. (Student # 1394, bilingual).

English is to increase my knowledge, to be able to read comics and novels, to understand songs in English, and to help me for the future. (Student # 1006, monolingual).

English teaches me to start a new life. It could help me. English is like a smile in my heart that gives blessing in my heart. (Student # 1141, monolingual).

I wish all Indonesian people to be able to speak English. Amen. (Student # 1183, monolingual).

At first learning English, it was boring enough, but gradually it turned out quite fun and enjoyable, because I started to understand it. Then?... I like it. (Student # 1303, monolingual).

My experience in learning English is that I am trying to enjoy the lessons even though I do not like English. I am trying to understand words in English. (Student # 1375, monolingual).

I often converse in English alone at home. I often speak with my younger siblings, for example: 'eat', 'drink', in English. (Student # 1360, bilingual).

When I was at kindergarten, I used to be afraid that I could not read and understand English. But as time goes, my knowledge of English builds up. At present I often use English in my daily life. (Student # 1363, bilingual).

English is fun but at the same time it is annoying. However, thinking deeply, I prefer English to my local language (Acehnese). It is easier to learn English than Acehnese. (Student # 1072, monolingual).

Yes, of course. When I was still poor in English, one overseas man asked me, "where do you live"? I was young at that time and could only answer, "I don't know", and shook hands with him confidently. (Student # 1296, monolingual).

Summarising this theme, the majority of both bilinguals and monolinguals stated that they liked learning English because it was fun and enjoyable, easy, important, unique (that is, pleasing on listen but difficult to pronounce), yet challenging. Some students also stated that dictionaries helped them. They were happy with their

vocabulary, they wanted to learn English with simple sentences, and they believed that was enjoyable, if studied correctly, but that needs time to learn.

Theme 5: Common Views on the Benefits of Learning English

Many students (bilinguals, N=77; monolinguals, N=66) stated that the benefit of learning English was that it could help them to go abroad. One hundred and thirty-five comments said that the students learnt English to help them in higher study (bilinguals, N=76; monolinguals, N=59); while 87 students stated that they needed to learn English because it was an international language (bilinguals, N=46; monolinguals, N=41), and 75 commented that English was a means to pursue dream jobs (bilinguals, N=46; monolinguals, N=27). Twenty-five comments showed that English helped them in their daily activities (bilinguals, N=16; monolinguals, N=9) and 25 comments revealed that English could help them study abroad (bilinguals, N=17; monolinguals, N=6). The rest stated that English was regarded as an additional language (N=10) (bilinguals, N=3; monolinguals, N=7); that it was used in the classroom (bilinguals, N=6); that it could help them to get enrolled in favourite schools (N=3) (bilinguals, N=2; monolinguals, N=1); and that it could help them to get a scholarship for higher studies (bilinguals, N=1) (see Figure 11.1).

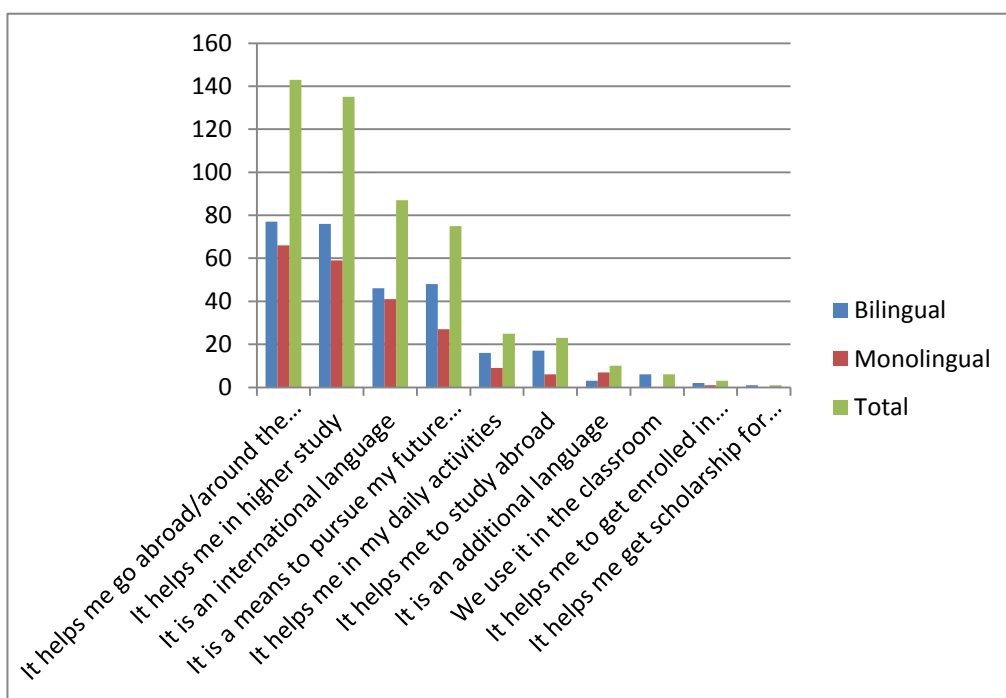


Figure 11.1 Students' Common Views on the Benefits of Learning English

Students' quotations:

Many students were very positive about learning English:

I learn English in order that I become a successful person and enable me to understand plenty of problems in English, and also to be able to work with important people overseas when I become a successful person. (Student # 1071, bilingual).

I like learning English because it could help me to study in Sydney. I really want to study medical sciences there. I have been learning English in many ways, such as listening to English songs; films; and story-books. I also learn English with the help of my Mum, because she is an English and German teacher. My favourite song is 'Hush Hush'. (Student # 1168, bilingual).

I like English, because with that I am free to go to any country with different languages (by using English), also I do not need to bother with the languages used in those countries. (Student # 1191, bilingual).

My experience in learning English is that I know English now. English is one of the international languages. If we go overseas, if we do not know English, we won't understand anything, and we cannot do anything. But, if we know English, we can easily understand what the people are talking about, and we can easily acquire whatever we need overseas. (Student # 1203, bilingual).

When I was in Singapore on holiday, I got lost. I wanted to ask the way to get back to my hotel, but they did not understand me because they used English. Fortunately, I know English a little bit. So by learning English we can understand it. (Student # 1207, bilingual).

English is so enjoyable. Especially in the globalisation era, it is a must for us to speak English fluently. Because nowadays, both Mathematics and English are the most important subjects, so automatically we have to be fluent in English. (Student # 1335, bilingual).

In my opinion, English is the coolest subject to learn because if we are able to master English, it will be needed and we can pursue our study overseas. (Student # 1393, bilingual).

I like English, because when I encounter some overseas people having a holiday in my city, I can speak to them. (Student # 1047, monolingual).

I really like English because I want to be a medical doctor working overseas. (Student # 1065, monolingual).

I like learning English, because if I can speak English, I can speak with foreigners when I am older, and also can travel around the world. If we can

Speak English it could help us get a good job, and we can even work overseas. (Student # 1279, monolingual).

In my experience, I am happy because with English I can speak with foreigners or can understand what they are talking about. (Student # 1345, monolingual).

One student was initially negative about English, but then changed

Actually I did not like English, but interestingly, now I start to like it, because if I can speak English I can pursue my higher study overseas. (Student # 1377, bilingual).

To sum up, most of the students learnt English because they believed that it would help them to go abroad, or around the world, and help them in higher study later. The students believed that English was worth studying because it is an international language and it is a means to pursue a dream job. Some other students liked English as an additional language to help them in their daily activities.

Theme 6: Confidence and Achievement in Learning English

Many students (bilinguals, N=151, and monolinguals, N=87) stated that they were motivated to learn English. Some students (bilinguals, N=18, and monolinguals, N=7) took after-school English tuition to improve their English. Other students said that they had won an English competition (bilinguals, N=11; monolinguals, N=7), were happy with their vocabulary (bilinguals, N=1; monolinguals, N=3), liked learning English from a very young age (bilinguals, N=11), and felt lucky to understand English (bilinguals, N=7).

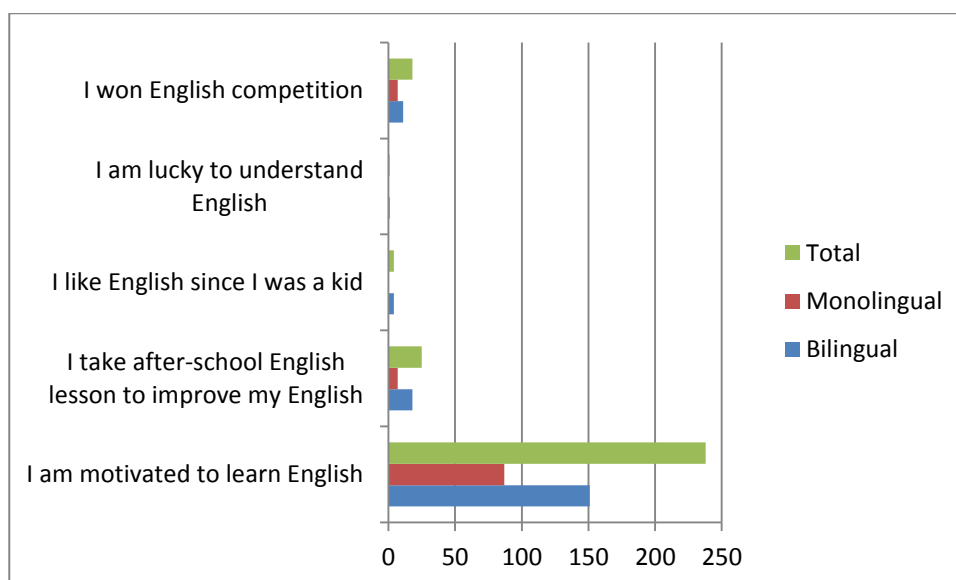


Figure 11.2 Students' Confidence and Achievement

Students' quotations:

I was so happy when taking part in a story-telling competition. I did not win anything at that time but my brother did. I was happy enough because my English teacher said that my pronunciation was good. I guess I have to study English harder. (Student # 1111, bilingual).

First time studying English I could only say, 'yes/no'. But from that day on I keep learning English. It turns out that English is fun. Moreover English is a famous language and is used daily. (Student # 1299, bilingual).

Yes. It took me quite some time until I understand English. Actually it does not work if we learn English only at school; therefore, I study it again at home with the help of my older sister and my parents. Gradually I understand it. (Student # 1354, bilingual).

I prefer when it has songs as well. I am happy with English and I am quite fluent in it. I am good at it now because I like to watch Japanese films with English subtitles. I also like western films. (Student # 1389, bilingual).

I like English so much; especially when I lead my friends reading some texts on the whiteboard, WOW... It feels like I am an honour person. I would like to be a professional English teacher when I grow up. (Student # 1057, monolingual).

I like English very much, because my first and prime wish is to be able to speak and write in English well, clearly and correctly. (Student # 1061, monolingual).

I have experience in learning English. At this age, I am able to speak English and use it within my school. However when I used it in my classroom, some friends said, “you show off with your English”. Probably they said that because they do not understand English. (Student # 1381, bilingual).

To summarise, the majority of respondents from both bilingually-taught and monolingually-taught schools responded that they were motivated to learn English. They took after-school lessons or tuition and some were happy because they had won some English competitions.

Theme 7: Learning English Through the Media

More bilinguals (N=27), commented that they liked playing games, singing songs, watching TV programs, and watching English films, either to understand, maintain or improve their English. The rest (N=9) stated that they were aware that number of media, software, and computer games used English. Therefore, they wanted to know English in order to use those media correctly. Of the comments, only one revealed that she or he preferred learning English through his or her school laboratory.

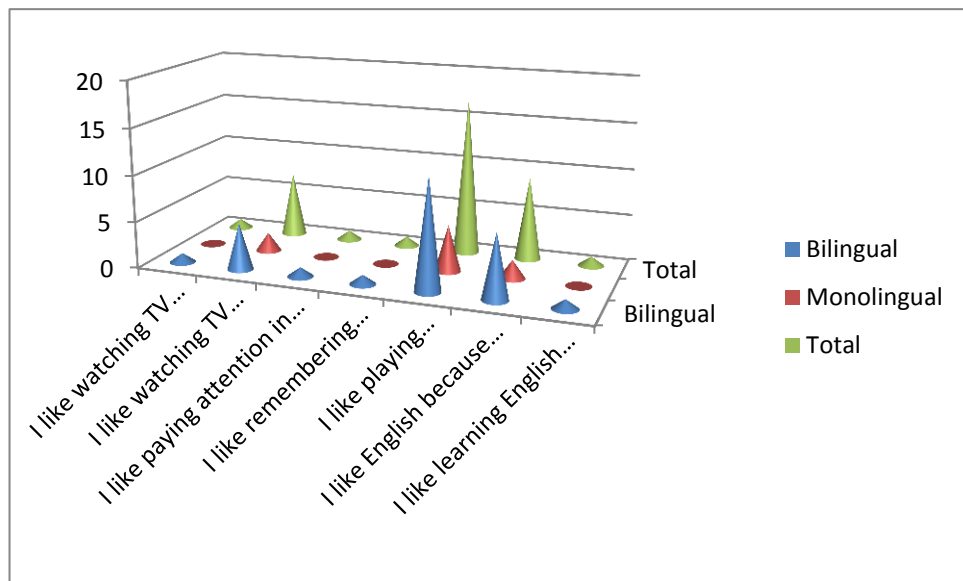


Figure 11.3 Learning English through the Media

Students' quotations:

Yes. Every Sunday and on public holidays I am learning through Australia Radio. The session has 104 lessons for 2 years. I learn English to ease my way to

go abroad and to study in the Netherlands or Australia. (Student # 1310, monolingual).

I like English because it is an international language and almost all games and software are in English. (Student # 1365, bilingual).

I like English for my future and in order to know the content of games at my house. (Student # 1378, monolingual).

I love English because it could help me to understand English in games as well as my daily life. (Student # 1382, monolingual).

I like English. I also like to listen to English songs and watch western movies. Learning English is a lot of fun. (Student # 1386, monolingual).

I like understanding English through English-based videos. (Student # 1031, monolingual).

I often watch western films with English subtitles when learning English. Should I find a new word, I would take a note, then refer to a dictionary. I would retell my English teacher the story on the following day. (Student # 1159, monolingual).

I often watch films and songs in English. It really helps me to increase my English vocabulary. (Student # 1175, monolingual).

I was once invited to talk with my overseas teacher at an English tuition. She/he also asked me to play games. She/he spoke so fast that I did not understand. (Student # 1183, monolingual).

To sum up, some students liked playing games in English, listening to English songs, and reading English story books in order to improve their English. They were also aware that numerous forms of media, software, games, computer programs, or videos use English.

Theme 8: Family Support in Learning English

There were 43 students who commented on family support (bilinguals, N=37; monolinguals, N=6). Their family, that is, fathers, mothers, uncles, aunts, and siblings encouraged them to learn English. Thirty-three bilinguals liked to practise English conversation with their older brothers or sisters. The students emphasised that they learnt English well in order to make their family proud. A few commented

(monolinguals, N=7) that one of the reasons they learnt English was because they wanted to teach English to their family members.

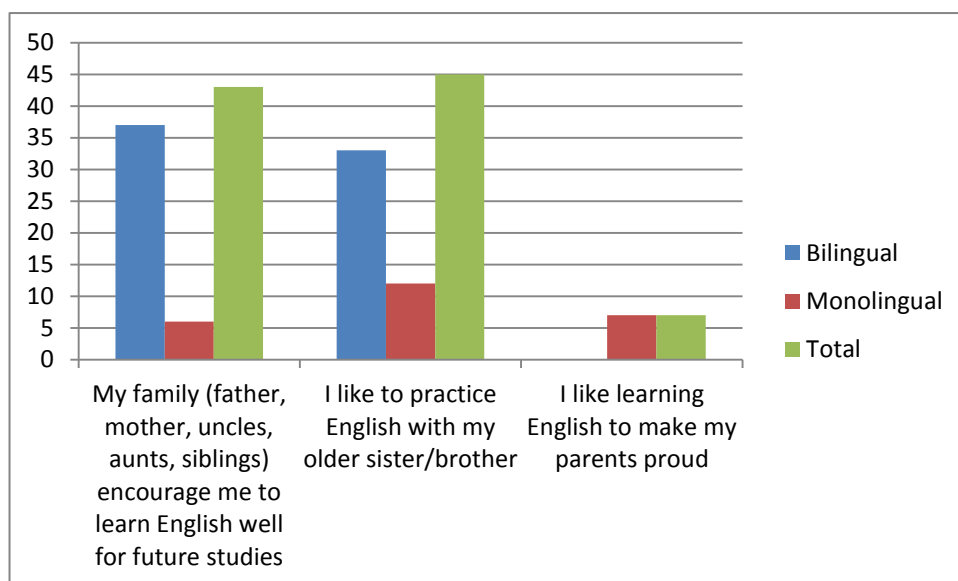


Figure 11.4 Family Support in Learning English

Students' quotations:

My parents often help me learning English, for example, my mum often ask me to communicate in English; my father often buys me English storybooks. A couple of my friends on Facebook are living overseas, and I love chatting with them in English. (Student # 1251, bilingual).

Yes. I did not like English. But, because my mother is an English teacher and I listen to her talking English with my bigger sister, then, I am interested in it. (Student # 1393, bilingual).

English can make us smart and successful, and can also make our parents proud. (Student # 1062, bilingual).

I like English very much because it could help me go overseas to be an English interpreter and can add some value to me. My parents will be so proud of me if I can understand English. (Student # 1349, bilingual).

I learn English with members of my family. It happens that my mum and my cousin know English a little bit. I like English because it is fun, and cool (Student # 1282, bilingual)

I love English very much. I have always got good scores. Thanks to my teacher and my elder sister who is studying to be an English teacher. I often ask her words/meanings that I do not know about. I also usually speak English with her. (Student # 1371, bilingual)

I like learning English with my teacher so much. I also learn it with my elder sister. (Student # 1102, monolingual)

It is easy to learn English as long as we learn it seriously. Also, we will be able to use it soon. English is the most interesting. (Student # 1113, monolingual)

When I was 4 years old, I got help from my mum, sister, and cousin (who was an English lecturer). I often seek help from her and she helps me much. (Student # 1151, monolingual)

I like English lesson because I was taught to speak English when I was 5 years old. Thanks to God, I love English so much now. It started from my mum. I got motivation from my family who are already fluent in English; therefore I have to be fluent in English, too. (Student # 1060, monolingual)

This theme summarises that some students received help from their parents and older siblings, as well as from other family members regarding the learning of English. They also liked to learn English in order to make their family proud of them.

Theme 9: Obstacles in Learning English

Some students claimed English was a difficult subject (bilinguals, N=31, Monolinguals, N=41). They said that vocabulary, meanings and grammar were not easy to understand (bilinguals, N=15, monolinguals, N=21). English was not easy to learn (bilinguals, N=6, monolinguals, N=22) and the four skills, that is listening, speaking, reading, and writing, were difficult to acquire (N=34 for both). Four students stated that they did not like English because English was not needed in Indonesia. Only eight students (bilinguals, N=3, monolinguals, N=5) wished to know English better; and seven students (bilinguals, N=5, monolinguals, N=2) were aware that they were not good at English.

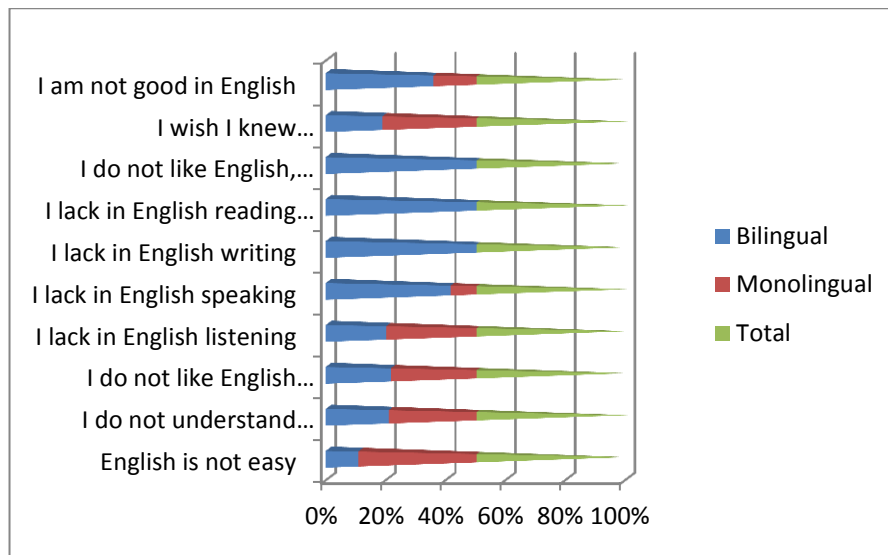


Figure 11.5 Obstacles in Learning English

Students’ quotations:

I like English, but I dislike it a bit, because when writing it, the words written are different to the way they are pronounced. I find it difficult when we do dictations with my English teachers. (Student # 1180, bilingual).

I love learning English because English can help me do higher study. But sometimes it is difficult. The most difficult ones are on ‘pronouns’, ‘possessive adjectives’, and ‘possessive pronouns’. (Student # 1313, bilingual).

My English is not fluent because I do not like English. It is so difficult for me, so I like sports instead. However English is also important to help me for future studies and when I am overseas. (Student # 1009, monolingual).

No. I do not like English much. (Student # 1012, monolingual).

In my opinion, learning English is fun, however, I do not know why I do not like it. (Student # 1284, bilingual).

There are plenty of grammar/rules in English. They make me so confused. (Student # 1292, monolingual).

Actually I like English, but sometimes I do not know why I do not understand what people are saying in English. I think I won’t be able to speak English even though I have been crossing the seas. Thank you. (Student # 1043, bilingual).

I do not like English because there are plenty of English words that I do not know. That sometimes make me find English annoying. Sometimes I do not know what people are talking about English are. (Student # 1148, bilingual).

I do not like English so much because it is difficult to understand. Besides, we do not communicate in English in our country. (Student # 1139, bilingual).

As a summary, English is viewed as a difficult additional language for some students, in terms of vocabulary, grammar, and meanings. A few students were aware that they lacked ability in listening, reading, and writing and others wished that they knew English better.

Theme 10: Other Views

Five students found that the tests conducted by the researcher were beneficial and wrote that the tests gave them good insights into English lessons. They were considered to be the first tests that were well prepared and presented.

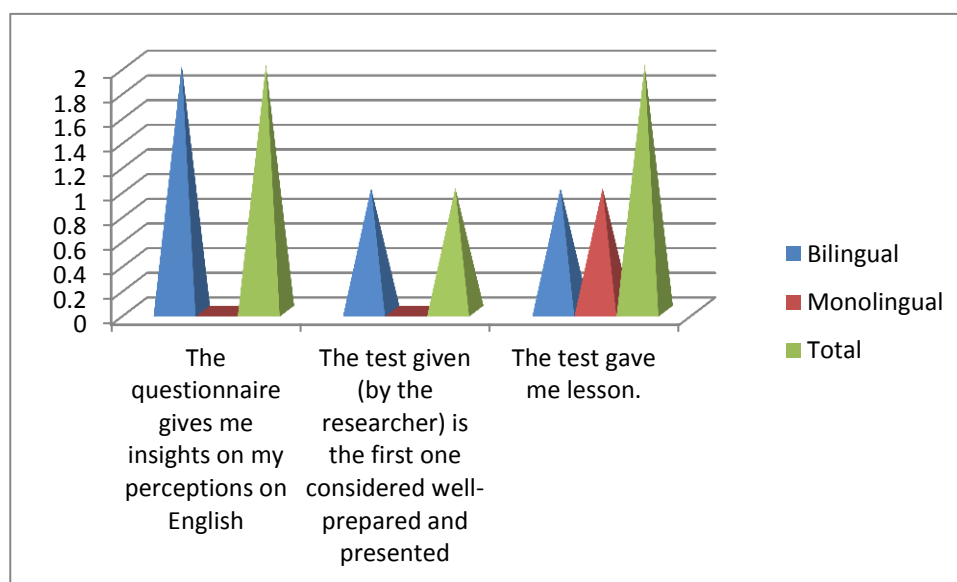


Figure 11.6 Other Views

Students' quotations:

I have been experiencing good English experiences when it was taught by Mr. R in this school. And this is the very first well-presented Reading Comprehension test that I have ever had. (Student # 1061, bilingual).

Thank you. With this questionnaire, I can tell my feelings, and through this questionnaire I know more than before. I like answering this questionnaire. (Student # 1077, bilingual).

Summary

Using the Miles and Huberman (1994) method, ten themes were created from 1110 comments from bilingual students and 732 comments from monolingual students. Both bilinguals and monolinguals shared similar opinions about the ten themes. The students' written comments generally expressed their positive attitudes with regard to learning English as a foreign language. Most of the students thought that learning English was important for their present and future needs. Even though they faced some obstacles in learning English, they kept themselves motivated, and stated that they wanted to continue to learn English as well as Indonesian.

The results of the qualitative analysis added more information to the results of the quantitative analysis. Both analyses provided new knowledge on Acehese students' abilities in English reading comprehension, English writing, and students' attitude and behaviour, as well as on the students' opinions and thoughts regarding their experiences in learning English as a foreign language.

The next chapter answers the research questions, and then discusses the results and the implications of the present research study.

CHAPTER TWELVE

DISCUSSION OF RESULTS AND IMPLICATIONS

This concluding chapter presents an overview of the study conducted, gives answers to the research questions, provides a summary of the findings, as well as presents discussion about bilingual education. It also provides some limitations and implications of the study and gives some recommendations for further research. The present study was aimed at investigating whether students who were learning English bilingually through the combined medium of English language and Bahasa Indonesia were better in some English abilities (English Reading Comprehension and English Writing) compared to students who were learning English with the medium of Bahasa Indonesia only. It also sought a comparison between students learning English bilingually and students learning English monolingually in their attitude and behaviour with regard to learning English as a foreign language.

I live in Banda Aceh and, from anecdotal evidence, I know that the perceptions and beliefs of the Acehnese community, consisting of parents, students, and teachers are that bilinguals at state schools perform significantly better than their counterparts. These perceptions and beliefs seemed to occur after some academic results at certain schools following the implementation of the new 2004 English curriculum that required every state school, from elementary level to secondary level, to provide some classes with bilingual programs.

In Aceh Province, the implementation of the new English curriculum was supported by a number of international workers who were welcomed to help reconstruct and rebuild Aceh after the 2004 Tsunami catastrophe. During that time some international organisations established bilingual boarding schools, where only English and Turkish were used. These schools soon gained in popularity.

Popularity was not only enjoyed by the Turkish boarding schools, but also state schools with bilingual programs. Large numbers of parents wanted to enrol their

children at the state schools with bilingual programs. Teachers also encouraged parents and student candidates to enrol at bilingual schools.

However, there was no known research evidence as far as could be ascertained, in Indonesia that bilinguals outperformed monolinguals. There was not a single research study reported and conducted in Aceh Province, or in any other provinces of Indonesia, that investigated this issue. Therefore, the present study was conducted to investigate the answer as to whether or not state school bilinguals did perform better than state school monolinguals in Aceh Province.

Summary of the Present Study

The data collection for the present study was carried out in Banda Aceh, from August 2010 until April 2011. The pilot study was carried out in October and November 2010, and the main data collection was administered from January 2011 until April 2011. Thirteen state Middle Schools took part in the study, with 780 female and male student respondents, consisting of 394 students from bilingually-taught schools (N=5) and 386 students from monolingually-taught schools (N=8). The study involved control and experimental groups with pretests and posttests using three measures: (1) English Reading Comprehension; (2) English Text Writing; and (3) an Attitude and Behaviour Questionnaire. All schools share similarity in terms of English curricula, English syllabi, English lesson content, English lesson times, English lesson duration, homework, and text books, which then were taken into account for recruitment of student-respondents of the two types.

The 780 students sat for three tests: English Reading Comprehension, English Writing; and Behaviour Questionnaire, for both pretests and posttests. The data analysis for three variables was conducted with the RUMM2030 computer program (Andrich, et al., 2010), to create linear scales, and a Mixed Design ANOVA and Two-Way ANCOVA (SPSS) were conducted to test for differences between the control and experimental groups.

This dissertation consists of twelve chapters. *Chapter One* started by introducing readers to the Indonesian educational system and to bilingual and monolingual education in Aceh province, Indonesia. It also presented the rationale and

background behind the study, as well as its aims. *Chapter Two* briefly revisited the nature of bilingualism, the nature of bilingual education, including bilingual education in Indonesia, and summarised some of the more relevant and recent research (2003-2013) on these topics from countries across the world. The conceptual framework of bilingual and monolingual teaching, measurement, and the variables of the study were presented in *Chapter Three*. The chapter informed readers about the conceptual framework behind bilingual and monolingual teaching in Aceh province, Indonesia, and explained why bilingual education was expected to produce superior achievements over monolingual education after a two month experiment. It also presented the three dependent variables of the study: English Reading Comprehension, English Writing, and Behaviour with regard to learning English, and explained the conceptual structure of each variable. Then there also was an explanation of the problems of True Score Theory measurement and why it was not used in the present study, but a better alternative – Rasch measurement – was used instead. *Chapter Four* explained the design of the study, covering the research strategy and design; mixed-method design; samples; pilot study; study and ethics approvals; control of extraneous variables in the quasi-experiments, and test data collection, data entry, and analysis of the quantitative and qualitative data.

The findings of the study were presented in Chapters Five to Eleven. *Chapter Five* showed the findings of the Rasch analysis of the Reading Comprehension. *Chapter Six* reported the experimental results on Reading Comprehension. The results of the Rasch analysis of English Writing were presented in *Chapter Seven*. The findings of the experimental results on English Writing were shown in *Chapter Eight*. *Chapter Nine* reported the findings of Rasch Measurement of Behaviour. The results of the experimental findings of the Behaviour were showed in *Chapter Ten*. *Chapter Eleven* reported students' written comments with regard to learning English as a foreign language. *Chapter Twelve*, the concluding chapter, presented a discussion of the results and the implications of the present study.

Answering the Research Questions

The purpose of the present study was three-fold. The first was to investigate the achievements of first year middle school students in Banda Aceh (Indonesia) in English

text writing, English reading comprehension, and behaviour with regard to learning English, as dependent variables, in the context of differences in gender and school types (bilingual/monolingual). The second was to investigate the behaviour of students who have been taught English bilingually and of those who have been taught English monolingually, in term of their ability in English. The third is to investigate the students' perceptions, through their written comments, in relation to their learning English as a foreign language.

The study's purposes were presented in four research questions, which guided the investigation and informed the methods used, as well as the analysis of the quantitative and qualitative data.

Research Question 1 *Do first year middle state-school bilinguals in Banda Aceh achieve better in English Reading comprehension and English text writing than those of monolinguals?*

Based on the Rasch-created linear scale for Reading Comprehension (see Chapter Five), the first year middle school bilinguals in Banda Aceh achieved better in English Reading Comprehension than monolinguals. There was a significant difference in output by bilinguals compared to monolinguals, with bilinguals achieving at a higher standard on most items. This is supported by a significant difference in overall Rasch measures for Reading Comprehension by type of teaching (either bilingually or monolingually), in favour of bilingual teaching.

The result for the Rasch analysis was supported by the experimental results using SPSS ANOVA and ANCOVA (summarised in Chapter Six). Bilinguals significantly outperformed monolinguals on English Reading Comprehension.

Results for English Writing were reported in Chapter Seven and Chapter Eight. Based on the Rasch-created linear scales, bilinguals had a statistically significantly better English Writing ability than monolinguals.

The results of the Rasch analysis (summarised in Chapter Seven) on English Writing was supported by the experimental results using SPSS ANOVA and ANCOVA

(Chapter Eight). Bilinguals performed significantly better than monolinguals on English Writing, with $F(1,775) = 124.69$, $p = 0,000$, with partial eta squared = 0.139.

Research conducted in countries other than Indonesia has supported the view that bilingual teaching produces superior results to monolingual teaching. A summary of research done over the last decade (2003-2013) supports this view.

Research Question 2 *Do first year middle state school bilingually-students at bilingual schools in Banda Aceh have better behaviours with regard to learning English than those at monolingual schools?*

Based on the results of the Rasch-created linear scale for Behaviour (Chapter Nine), first year middle school bilinguals in Banda Aceh had better behaviours with regard to learning English than monolinguals. This is supported by the difference in overall Rasch measures for Behaviour by type of teaching, which was statistically significant in favour of bilingual teaching.

The results of the experimental study were in agreement with this. There was a main effect between bilinguals and monolinguals, where bilinguals had better behaviours with regard to learning English as a foreign language.

Research Question 3 *What are the attitudes and behaviours of first-year middle state school bilinguals with regard to bilingual and monolingual education in Banda Aceh in term of learning English?*

A Rasch-created linear scale of Attitude and Behaviour with regard to learning English as a foreign language provided a list of behaviours ranging from the easiest for the students to perform, to the most difficult. They are listed here from easy to hard:

1. I actually like the way my teacher teaches English writing. (*easiest*)
2. I actually pay attention to someone speaking English.
3. I actually read words in English carefully.
4. I wish to say new English words several times.
5. I actually can understand English better when I do activities with friends.
6. I actually do read English at home on my own.
7. I actually learn more when I study English in groups.
8. I actually like English because we use it in the classroom.
9. I actually practise English with other students.
10. I actually say new English words several times.
11. I actually start conversations in English with my friends. (*hardest*)

The ten statements from the Behaviour and one statement from the Attitude list have been categorised into three levels of difficulty, based on the students' answers. Statement numbers 1-4 are returned as easy, 5-8 as quite difficult, and 9-11 as most difficult. Statements numbers 1-4, such as statement number (1) *I actually like the way my teacher teaches English writing*, fell under the category of 'easy' due to the fact that they were answered positively by the majority of the students. As the present study did not investigate the reasons behind the students' answers, it was assumed that the reason for answering those statements positively was because the activities involved students' personal willingness, which seemed easier to perform than statements with activities that involved other people (numbers 5-8), such as statement numbers (5) *I actually understand English better when I do activities with friends*, and (7) *I actually learn more when I study English in groups*.

Statements numbers 9-11 were considered difficult. There was only a small number of students who answered those statements positively, which was assumed to be because the stated activities involved active engagement in English, such as statement numbers (9) *I actually practise English with other students*, and (11) *I actually start conversations in English with my friends*.

Research Question 4 *What are first-year middle state school students' perceptions with regard to their learning of English as a foreign language?*

Results from the qualitative data (reported in Chapter Eleven) revealed that the students had positive perceptions towards their learning of English as a foreign language. The students' perceptions were categorised into ten themes. The following are the themes with statements the students mostly wrote for each theme.

Theme 1: Tasks on Learning English as a foreign language

- I like listening to English songs
- I like speaking in English
- I like learning English through reading comprehension
- I like writing

Theme 2: Student-Student Relationship

- I learn more when I study English in groups

Theme 3: Student-Teacher Relationships

- I like the way my teacher teaches English

Theme 4: Personal Views of Learning English

- I like learning English

Theme 5: Common Views on the Benefits of Learning English

- I like English because it helps me go abroad/around the world

Theme 6: Students' Confidence and Achievement in Learning English

- I am motivated to learn English

Theme 7: Learning English through the Media

- I like playing games/singing songs/reading story-books to improve my English

Theme 8: Family Support in Learning English

- I like to practise English with my older sister/brother

Theme 9: Obstacles in Learning English

- I do not like English because it is not easy to learn

Theme 10: Other Views

- The questionnaire gives me insights into my perceptions of English

Amongst the ten themes, the most frequently mentioned comments related to ‘Common Views on the Benefits of Learning English’, with 505 written comments, followed by ‘Personal Views of Learning English’ with 408 written comments, and ‘Students’ Confidence and Achievement in Learning English’, with 286 written comments. The least frequently mentioned themes were ‘Learning English through the Media’, ‘Student-Teacher Relationships’ and, ‘Other Views’, with 37, 29, and 10 written comments respectively.

Generally, the students’ written comments showed that they were very positive about learning English in relation to the way they learnt and experienced English. They liked learning tasks regarding the four English skills (listening, speaking, reading comprehension, and writing). They mentioned that they had good relationships with their fellow students and teachers. Even though some students faced difficulties in learning English due to its complexity in vocabulary, grammar, and meanings, most students revealed that they liked learning English because it was challenging and enjoyable, as well as helping them to go abroad and helping them in higher education.

The students also stated that they were motivated to learn English. Some students took after-school English lessons to improve their English, as well as learning through the media. In addition to this, they received good support and attention from their family. Some students mentioned that the tests that the researcher gave them helped give them insights into their perceptions of English.

Summary of the Findings

The main findings are set out under: (1) English Reading Comprehension, (2) English Writing, and (3) Questionnaire.

A. English Reading Comprehension:

1. Both the bilinguals and monolinguals in Banda Aceh achieved better scores in English Reading Comprehension posttests than in the pretests.
2. Bilinguals performed better in the pretest than monolinguals.
3. Bilinguals performed better in the posttest than monolinguals.
4. Female bilinguals performed better than male bilinguals.
5. Female monolinguals performed better than male monolinguals.

6. Female bilinguals performed better than female monolinguals.
7. Male bilinguals performed better than male monolinguals.
8. Generally, female students outperformed male students.
9. Generally, bilinguals outperformed monolinguals.

B. English Writing:

Some caution should be exercised in interpreting the English Writing results since two months between pretest and posttest measures maybe too short to obtain strong improvements.

1. Bilinguals performed better in the English Writing pretest than monolinguals.
2. Bilinguals performed better in the English Writing posttest than monolinguals.
3. Female bilinguals performed better than male bilinguals in English Writing.
4. Female monolinguals performed better than male monolinguals in English Writing.
5. Female bilinguals performed better than female monolinguals on English Writing.
6. Male bilinguals performed better than male monolinguals in English Writing.
7. Generally, female students outperformed male students in English Writing.
8. The posttest results for English writing were measured to be lower, after two months, than the pretest result, and this is true for both females and males, and for both bilinguals and monolinguals.
9. In general, bilinguals outperformed monolinguals in English Writing.

C. Behaviour with regard to Learning English as a Foreign Language:

1. Bilinguals scored higher than monolinguals in the pretest Behaviour measure.
2. Bilinguals scored higher than monolinguals in the posttest Behaviour measure.
3. Female bilinguals scored higher than male bilinguals in the Behaviour measure.
4. Female monolinguals scored higher than male monolinguals in the Behaviour measure.
5. Female bilinguals scored higher than female monolinguals in the Behaviour measure.
6. Male bilinguals scored higher than male monolinguals in the Behaviour measure.

7. Generally, female students outperformed male students in the Behaviour measure.
8. In general, bilinguals outperformed monolinguals in the Behaviour measure.

Discussion about Bilingual Education

The benefits of bilingualism over monolingualism have been supported in many studies. The benefits have been associated with the enhanced cognition, attention control, and brain functioning in bilinguals. The findings of the present study support a number of previous bilingual studies regarding cognition and attention which have been conducted within the last decade (2003-2013) and support the anecdotal evidence from Aceh that children who were taught English bilingually at state schools outperform children who were taught English monolingually at state schools.

The work of Foy and Mann (2013) revealed that bilingual children outdid monolingual children in nonverbal auditory executive function tasks. Bilingual children significantly outperformed monolingual children on a series of standardised working memory, cognitive control, metalinguistic awareness, and problem-solving ability tests (Lauchlan, Parisi, & Fadda, 2012), and managed cross-linguistic interference more effectively than monolinguals (Bartolotti & Marian, 2012).

Further examples from the recent literature on bilinguals demonstrated higher levels of performance in vocabulary and phonological tasks (Kaushanskaya, Blumenfeld, & Marian, 2011). Bilingual students outperformed monolingual students in reading proficiency (Rauch, Naumann, & Jude, 2011). Bilingual children have advantages over their counterparts in regard to selecting phonological skills when language use and proficiency are controlled (Goldstein & Bunta, 2011). Bilinguals evidenced greater phonological complexity in Spanish than English with respect to word density (Freedman & Barlow, 2011). Bilingual students also scored higher than the monolingual children on a grammatical judgment test (Foursha-Stevenson & Nicoladis, 2011).

Bilingual college students performed better in a task-switching paradigm compared to monolingual college students (Prior & MacWhinney, 2009). Bilinguals were more efficient than monolinguals at inhibiting distracting information (Treccani,

Argyri, Sorace, & Salla, 2009). Bilinguals acquired a newly-learned language faster than monolinguals (Kaushanskaya & Marian, 2009). Bilinguals developed a better ability to maintain action goals to use for bias goal-related information (Colzato et al., 2008). Bilingual participants were faster in performing the task-switching paradigm as well more efficient in the alerting and executive control networks (Costa, Hernandez, & Sebastian-Galles, 2008). Bilingual children maintained better accuracy than monolingual children in dual-task paradigms, especially on visual tasks. Bilingual participants performed better in dual-task letter number category tests (Bialystok, Craik, & Ruocco, 2006). Bilinguals also achieved higher scores than monolinguals in the Simon Task (Bialystok, et al, 2004). Bilinguals had advantages over monolinguals in justifying their evaluation on the appropriateness of certain request strategies in particular contexts as well as on their use of request realisations (Jorda, 2003).

Being bilingualism for Indonesians, more specifically, Acehnese students, is seen as a privilege. There is nothing wrong about learning English, alongside learning Indonesian, despite the ban of bilingual education in Indonesia recently (2013). Learning English is believed to be important because English is now an international language and it is the language of information and technology. Without having ability in English, it is argued that Indonesian students could be predicted to be 'left behind'.

The importance of learning English was mentioned by the first year students of state middle schools in Banda Aceh. Most of them stated that they learnt English because they believed that English could help them go overseas, either for holidays or to pursue their further study. It is convenient for people to travel to neighbouring countries of Indonesia, for example, Singapore, Papua New Guinea, and Australia, if they have a good command in English. They, at least, can read signs at airports and cities or can manage simple conversations with the countries' people, for directions, or simply to get new friends. It is hard to imagine that, if somebody travels overseas without any ability in English, then they would have no trouble understanding signs, attractions, destinations and directions, without some help.

People who want to study overseas are now required to be able to have a good command of English. That requirement is not only for those who want to study at an English-speaking country but, it is also true for students who want to study in a country

which the language is similar to their own. Consider the example of Indonesian students who want to study in Malaysia: even though the Malay language and the Indonesian language are similar, if not identical in some words, and English is not their national language, the students need to pass an English requirement, such as TOEFL and IELTS. The reason behind that is because there are a great number of academic references which are written in English and whose equivalences are scarcely written in or translated into Malay. The case would be true for these students if they want to enroll in any Australian colleges or universities. In this country, English is the language for both academic and daily life. For those who seek a scholarship, their ability in English even is needed to be proved much earlier, that is when they need to satisfy the scholarship panel about their English ability in order that a scholarship be awarded to them.

Some students also mentioned that English could help them get enrolled in their favoured schools and help them get their dream jobs. Now, more favoured schools in Indonesia, especially in metropolitan cities, use English as a medium of instruction. That becomes a trend that attracts candidates of students to enroll at those certain schools. The schools usually run an entry test in order to ensure that only students with a good command of English are selected. Among the selection criteria is the ability for the students to speak, write, and read in English. Students with a low command of English could not go through. Similarly, regarding job vacancies, having a good command in English can give a credit to an employee candidate when competing for a good job. It is more likely that among some employee candidates who have similar academic achievement, age, and ability regarding that specific job, employee candidates with a good command of English would be favoured to get the job.

English is also a language for information and technology. Indonesian students are also aware that a great number of books are written in English. If they want to know the content of the books, they have to know English. In the same line, most technology comes in English. Most of games, smart phones, Facebook, Twitter, and the like, are programmed in English. Most of the manual books and booklets of television, refrigerators, air conditioners, radio cassettes and similar products, are written in English. One may find difficulty in operating them in Indonesia without ability in English.

Based on the discussion above, it establishes that there is nothing wrong about learning English by Indonesians, more specifically, Acehese students. Having more languages is better than having one language. Being bilingual is better than being monolingual. By being bilingual, his or her mind is more powerful than being monolingual. Although bilinguals are not necessarily smarter, they can be more flexible and resourceful. Bilingual learners are better at reasoning, at multitasking, at grasping, and at reconciling conflicting ideas (Kluger, 2013). So it is in the best interests of Indonesian students to be bilingual or multilingual.

Those skills start in the womb, continue at birth, crest at nine months, and decline at six years old. In the uterus, the third trimester is the first time for babies to be able to hear sounds. The first sounds for them to try to recognise are their mother language rhythms. When they are born, the babies recognise their mother tongue and distinguish it from other languages. If their mothers are bilingual, babies recognise both languages. The peak is at nine months when the babies have sharp ears for languages. By one year old, the door to languages begins to close. When the babies are about six years old, they are less natural to languages than they were when they were babies, but they are still better than teenagers or adults (Kluger, 2013).

The question now is whether to be bilingual or multilingual, or be worse off. If one wants to use it, Kluger (2013) suggests that one would better start learning it early, and maintaining it through a lifetime. Maintaining it is not an easy task. Considering a case of two siblings: 11 years and 6 years old who are living in Australia for four years. They have been learning English for four years and at the moment they have reached a resemblance to native Australians. Along the time, they also learn Bahasa Indonesia and Acehese language at home because those two languages are the languages of their parents. Upon returning to Indonesia after these four years, it is reasonably predictable that their ability in Australian English will fade, replaced by Bahasa Indonesia and Acehese language as the languages of the new community. Should they be able to maintain those three languages in a fair share, they might benefit more from the power of the bilingual and multilingual brain. The bilingual and multilingual brain, when activated from birth, can provide advantages for a lifetime, and even can reduce certain cognitive diseases (such as dementia) for old bilinguals and multilinguals to produce a lifetime skill (Kluger, 2013). Figure 12.1 shows bilingualism as a life time skill.

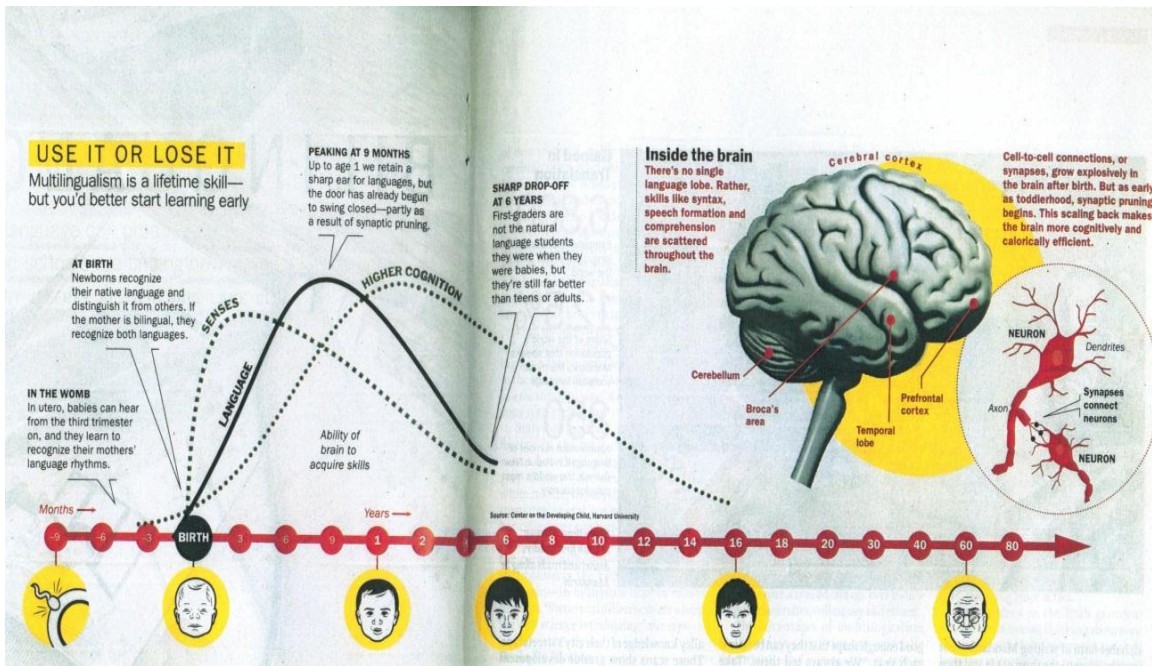


Figure 12.1 Bilingualism as a Life Time Skill

Source: Kluger, J. (2013). The power of the bilingual brain. *Time*, 182(5), pp. 26-27.

The results from Figure 12.1 have important implications for Indonesians in being bilingual or multilingual. It is in the best interests of all Indonesians, in both the long and short term, for bilingual education to be re-instated and to be available to all Indonesian students.

That female students do better at English Comprehension, English Writing and Behaviour with regards to Learning English is probably not a strong concern. In western countries this has been true for many years but males tend to catch up in late secondary school (college) and university. So it is probably not a concern for young Indonesian students as it can reasonably be expected that older male Indonesian students will also catch up in college and university years.

Implications for Policy and Practice

The present study has a number of implications for policy: for the Indonesian Ministry of Education; and for practice, relating to administrators, lecturers, trainee student teachers, Headmasters, ESL/EFL teachers, ESL/EFL students, and parents.

Implications for Policy: Indonesian Ministry of Education/Policy Makers

The implementation of the 2013 English curriculum in Indonesia that banned the bilingual programs from primary level until secondary level has been controversial. Some have seen it as a drawback with regard to learning English in Indonesia. Some have perceived it as a regretful decision, claiming that the decision made to ban bilingual education was immature, lacking in comprehension and lacking in research-based reasons.

The findings of the present study revealed that students who were taught English with two languages performed better than those who were taught English with Bahasa Indonesia only. They also showed that students who were taught bilingually gained better behaviour with regard to learning English. Regarding the implementation of the 2013 curriculum, the findings of the present study could contribute to giving a better understanding to the Indonesian Ministry of Education and its Policy Makers that bilingual education practices could not be blamed for the banning of bilingual education in Indonesia. As there could be a number of other factors with regard to what has been claimed as the 'unfairness of bilingual education in Indonesia', banning it may not be seen as a wise choice. Other factors, like government and school readiness in terms of financial support, human resources, and the like, should be taken into account as well. If the practice of bilingual education in Indonesia was seen as giving a further advantage to those who already have advantages, it should not need to lead to banning it, but to the government implementing bilingual education in all schools. By doing so, the government can maintain bilingual education and its advantages, and at the same time is able to control negative aspects that could harm the practice of bilingual education.

Implications for Practice

Implications for Administrators, Lecturers, and Trainee Student Teachers

The present study showed that teaching English to state middle school students using two languages (English and Bahasa Indonesia) enhanced students' English ability, especially in English Reading Comprehension, and English Writing, as well as improving their behaviour and their perceptions with regard to their learning of English as a foreign language. This has implications for administrators, lecturers, and ESL trainee-student teachers. At the administration level, the study suggested that a well-

planned bilingual education should be prioritised prior to any implementation of bilingual education in Indonesia. The situation in Indonesia tends to show that an ill-planned bilingual education caused the banning of the bilingual program throughout Indonesia. Bilingual programs that were implemented in 2004 were banned and the ban took effect in early 2013. Administrators should base their beliefs on research evidence and not rush to hasty decisions without regard to the evidence.

Previous experience with bilingual education implementation showed Indonesians that bilingual education had been improperly conducted. This led to unsatisfactory results claimed by certain groups who believed that bilingual education in Indonesia created unfairness toward students with low-income and from low socio-economic families (Revianur, 2013b). Further, these certain groups stated that bilingual education could only be enjoyed by students from high-income families who were able to support the schools financially. On the other hand, students from families with low incomes could only manage seats at schools without bilingual programs, creating an unfair perception. Groups of these unsatisfied parents actually brought the case to the Judicial Court (called *Mahkamah Konstitusi* in Indonesia). It resulted in the Judicial Court banning bilingual education in Indonesia from the 2013 (Revianur, 2013b).

However, a better future for bilingual education could be predicted, based on research evidence. The present 2013 English curriculum might last for five years. As Indonesia is dynamic and prone to change, it is highly anticipated that a new curriculum with bilingual programs might be reassigned after five years. If this prediction takes place, it is expected that Indonesian administrators will prepare a new bilingual program prior to its implementation. This is regarded as highly important in order to achieve a 'fairness of Indonesian bilingual education', which is fair and mandatory for each and every student in Indonesia. Therefore, it is hoped that there will be no more banning of bilingual education in the history of Indonesian education.

In line with this, lecturers and teacher trainers at universities should provide a proper and successful transfer of ESL/EFL knowledge, skills, and content to ESL/EFL trainee-student teachers and it should be compulsory, since there are clear benefits to all Indonesians. Trainee-student teachers need to be ready to teach their future students bilingually. One cannot expect well-taught bilingual classrooms if the teachers

themselves are not ready to teach them with knowledge and skills which are changing and demanding (O'Neill, 2008).

Implications for Headmasters, ESL/EFL Teachers, and ESL/EFL Students.

The present study gave evidence that bilingual education enhanced students' ability in English subjects, as it does in many western countries. This information is of importance to headmasters, ESL/EFL teachers, and ESL/EFL students, not only in Aceh Province but also in Indonesia as a whole. Being well-informed about the benefits of bilingual education in Indonesia, headmasters could play crucial roles in managing to prepare for future bilingual education support, in terms of financial support, human resources, administration resources, libraries, and other related school-maintenance support. Headmasters and others need to have access to research reports worldwide showing the benefits of bilingual education for all students and older citizens.

ESL/EFL teachers, in line with this, could try to do their best to implement a bilingual program in their classrooms by conducting teaching and learning processes bilingually, as well as assessing students' work consistently and fairly. They need to constantly upgrade their knowledge and proficiency in English in order to keep them updated with current English practices, techniques, approaches, and skills.

ESL/EFL students can enjoy the learning by using two languages (Bahasa Indonesia and English). Students can succeed in their schools well and acquire high motivation in learning English as a foreign language, but Headmasters need the research evidence to be widely available in order to be able to argue for bilingual education. Students can also use their previous knowledge to help them understand and improve their current knowledge of English. In addition to that, they can benefit from their exposure to English outside their classrooms, as English can also be learnt from books written in English that students read at home or in the library, or from TV programs and music to which they watch or listen, or conversations exchanged with tourists, or on social media like Facebook and Twitter, and the like.

Implication for the Banning of Bilingual Education

The present study could inform the education policy makers about the current reality in Aceh that bilingual education was still needed, apart from the banning of bilingual education throughout Indonesia. It also suggests that the Acehnese government should reconsider the banning of bilingual education in Aceh Province and reestablish bilingual programs at state schools in Aceh with the power it already has as a province. It has the privilege to manage its own education policy, and to learnt from past lessons and previous experience in regard to bilingual education.

Implication for Classroom Practices

The present study also provided implications for success in bilingually-taught classrooms. Teachers of state middle schools in Aceh could take the benefits from this study (and other studies) in order to become well informed about criteria on running a bilingual program, from practices conducted all over the world. The present study also presented examples of well organised English Reading Comprehension and English Writing tests, as well as a good questionnaire to make linear measures. These examples are worth trying in teachers' own classrooms. From some brief informal contacts made by the researcher with some of the respondents, it was revealed that some students have never experienced having Reading Comprehension, English Writing, and Questionnaire tests of the high quality used in this research. The students liked the paper test presentation, due to its clear instructions and interesting pictures that helped them understand the test more easily. This probably means that more money needs to be provided by government and policy makers to enable teachers to have access to high quality material. It also probably means that teachers need better in-service courses that provide research evidence for the benefits of bilingual education.

In practice in Banda Aceh state schools, English test presentations have not always been clear. Instructions have been mostly unclear and always needed further explanation, usually done by the test invigilators on the day of the tests. Pictures, especially colourful pictures, were seldom provided, due to high costs. With the example of the present study's paper test, findings implied that content and presentation of a test are considered equally important, in order that students perform their best.

Implications for Parents

There would seem to be a need to inform parents, through government circles and schools (parent/teacher meetings) about the benefits of bilingual teaching for students with immediate effects for young people and later effects for older people. There are benefits economically and for the young people of Indonesia who are fluent in English and Indonesian. Young Indonesians would be equipped to travel overseas to gain the skills and knowledge that are needed to improve their country. Especially nowadays, the Indonesian government provides a large number of scholarships for eligible candidates to pursue studies overseas, and English mastery is part of the criteria for the selection.

Implications for Further Research

Improved Rasch Measures

In the present study, the researcher designed and constructed novel test formats for three variables: (1) English Reading Comprehension, (2) English Writing, and (3) an Attitude and Behaviour Questionnaire. The three kinds of tests were piloted prior to the main data collection. However, data for two of the three measures (English Reading Comprehension and English Writing) did not show an ideal fit to the Rasch measurement model. This was probably caused by such factors as the difference between student ability and the level of test difficulty and because the students at any one level of ability did not agree about the difficulties of all the items. Future research, therefore, is suggested to examine these issues and perhaps find a model to design (or create) items for English Reading Comprehension and English Writing. It is most probable that a qualitative study would be required where the researcher could talk to and question the students about their views on the test items after the students have answered the items.

It is probable too that, because there would likely be a large difference in ability between pretests and posttests, different tests might be required for the pretests and posttests that could be equated using Rasch measurement, perhaps with some overlapping items.

More Research on Bilingualism and Bilingual Education in Indonesia

There is a scarcity of research on bilingualism and bilingual education in Indonesia and in Aceh province, and it is suggested that more research on these topics should be conducted. It is also important that the research be accessible to Indonesian educators, the Indonesian people and other researchers. There is a possibility that there is some research on bilingualism and bilingual education in Indonesia or in the Aceh province but that is not readily accessible to Indonesians and Indonesian educators.

Inaccessible research and scarcity of research can cause some missing links in regards to future research. This can contribute to poor debating and poor decision-making in Indonesia (as in the recent court case involving the banning of bilingualism in Aceh schools). Therefore, it is important that studies on bilingualism and bilingual education in Indonesia and in Aceh be continuously conducted and the results be made accessible to educators and the public.

Longitudinal Research

Studies on bilingualism and bilingual education involve conducting research over time, so as to allow researchers to make repeated measures that are directly comparable. Longitudinal research is suitable for use in the studies of bilinguals and bilingual education, where the method can investigate developmental trends and comparisons between bilingual students and monolingual students, as well as comparisons among bilingual students themselves over short and long periods of time.

Most longitudinal studies employ observations or measures compared over various time intervals. A large problem here is controlling for extraneous variables and, the longer the time frame, the more likely that extraneous variables (like all the different things that different students do outside school hours), will confound the results. The present study involved a short time frame but there is a clear need to have some longer studies, perhaps over six or twelve months. However, this can be costly and difficult to manage. It is probable that a large amount of joint university researcher and teacher cooperation, and perhaps parent cooperation, will be needed. Also, pretest and posttest equating may be needed using the latest Rasch measurement techniques to link pretests with posttests that may contain different items.

Teachers and parents

Studies on bilingualism and bilingual education can cover a wide range of people because many people (teachers, administrators, parents and policy-makers) are involved in children's education. Teachers and parents are believed to have the most influence and power in teaching bilingualism and in promoting academic achievement. Research on teachers' and parents' influences in relation to students' bilingualism and bilingual education in Aceh is inadequate. So, future research should consider teachers and parents as the focus of the studies on bilingualism and bilingual education in Indonesia and Aceh province. This could involve a variety of approaches including control and experimental groups with pretest and posttest Rasch created linear measures, longitudinal studies and qualitative studies. As already mentioned, controlling for extraneous variables will be a problem and researchers will probably have to use a mixture of physical and statistical controls with more accurate Rasch measures to be able to make better conclusions from the data.

English listening and speaking skills

The present study focused on two English skills: Reading Comprehension and Writing but it did not investigate Listening and Speaking skills. Future research is encouraged to investigate students' English abilities in listening and speaking because these are also important English skills. Future research results could be combined with the present study's findings (repeated with improved Rasch measures) in various longitudinal and qualitative studies in Aceh. Altogether, a better view would be gained of students' abilities and achievement in the four skills related to learning English as a foreign language in the Acehnese context.

Limitations of the Study

This study, inevitably, had some shortcomings, in terms of a number of aspects:

Managing Time and Extraneous issues

The experimental study was conducted within two months, which is equivalent to 16 meetings. This short duration was due to three reasons: (1) to allow sufficient meetings for students to experience the two medium of instruction (bilingual mode and

monolingual mode), (2) to avoid inconvenience for teachers, and (3) to manage the control of extraneous variables.

The duration of 16 meetings was predicted to be sufficient for students to be taught either bilingually or monolingually in order to allow them to experience their English learning progress. It was also expected that at the end of the period, the students would be ready to sit for the posttests with expected changes in terms of their ability in English Reading Comprehension, English Writing, and the Attitude/Behaviour Questionnaire with regard to their learning English as a foreign language. The duration of two months was also seen as a convenient time allowance for the teachers to participate. It was strongly predicted that beyond two months, teachers might feel disturbed due to their other responsibilities and commitments to the schools and family. The duration was also limited in order to control for extraneous variables related to both types of students involved in the study. Both bilinguals and monolinguals needed to maintain similarities in a number of aspects, such as the same English lesson content, the same number of lessons and lesson durations, the same pretest and posttest duration, the same homework, and the same text books, but the nature of the bilingual and monolingual teaching program needed to be left unchanged. The only difference between the two groups that the study intended to maintain was that the one group experienced bilingual teaching and learning processes, while the other group experienced monolingual teaching and learning processes. This similarity needed to be maintained in order that any changes that appeared after the time frame of two months could be associated with the change due to the type of teaching instruction.

The changes, measured in this current study, therefore, suggest that bilinguals were better than monolinguals in the three tests: English Reading Comprehension, English Writing and Behaviour Questionnaire, due to the fact that bilinguals were taught using both English and Bahasa Indonesia, while monolinguals were taught using only Bahasa Indonesia as the medium of instruction.

However, the present study would be more reliable if it extended the experimental duration, that is, more than two months, providing extraneous variables could be sufficiently well controlled. This assumes that the longer duration between pretest and protest would show that students could demonstrate better ability in the

posttest. This improved ability could be associated with the use of English and Bahasa Indonesia (bilingually) in teaching English as a subject in the Acehese context. The study would be more comprehensive if it were conducted over three school years, starting at year 7 and concluding at year 9.

Managing Teacher Perceptions (through Interviews)

The present study was conducted with quantitative data, with a small amount on qualitative data. For future research it is suggested that, rather than just students, the study also collect data on the perceptions of teachers. Teachers' perceptions could enrich our understanding of their experiences and expectations regarding their own way of teaching English, as well as their assumptions and hopes for their students with regard to learning English as a foreign language.

Managing Classroom Observations

The present study did not investigate the way a classroom of teaching and learning was run, either. Therefore, it is recommended that future research could examine classroom observations of bilingual and monolingual teaching. Classroom observation data offer a valuable understanding and knowledge of how bilingual students interact in a classroom, compared to monolinguals, during the teaching and learning process.

Measures were not Ideal and Need Improvement

The present study used data collected from 780 students in 18 state middle Schools in Banda Aceh, with both males and females, and bilinguals and monolinguals. The test designs for the three measures (English Reading Comprehension, English Writing, and Behaviour Questionnaire) were especially constructed by the researcher due to the fact that there were no existing tests that were suitable for the present study. The three tests were piloted prior to collecting the main data using a class of bilinguals and a class of monolinguals. Because True Score Theory of Measurement (like percentage scores) – the measurement method universally used in schools – is non-linear, a modern method of measurement – Rasch measurement – was used in the present study to create linear and more reliable measures. While these Rasch measures were better than the percentage scores used by teachers, they were not as good as the

researcher would have liked because the fit to the measurement model was not ideal. However, many significant results were still able to be obtained. For future use of these tests, it is recommended that the tests be modified and amended.

There is a need for high quality tests of English writing and English comprehension that would be applicable to many different standards of English and so much further research is needed to produce them, preferably using Rasch measurement techniques. Alternatively, there is a need for a theoretical method that enables researchers to design tests from first principles that will produce data that fit a Rasch measurement model without any misfit. That could well be a better outcome, if someone could create such a model.

REFERENCES

- Abu Rabia, S., & Siegel, L. S. (2011). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of Psycholinguistic Research*, 31(6), 661-678.
- ADB, A. D. B. (2009). Aceh-Nias rehabilitation and reconstruction, from <http://www.adb.org/Documents/Produced-Under-TA/39127/39127-01-INO-DPTA.pdf>.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlation of bilingualism. *Review of Educational Research*, 80(2), 207-245.
- Adi-Japha, E., Berberich-Artzi, J., & Lidnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development Psychology*, 46, 93-105.
- Alarcon, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350.
- Ali, Z., Mukundan, J., Baki, R., & Ayub, A. F. M. (2012). Second language learners' attitude towards the methods of learning vocabulary. *English Language Teaching*, 5(4), 24-36.
- Alshumaimeri, Y. A., & Almasri, M. M. (2012). The effect of using WebQuest on reading comprehension performance of Saudi EFL students. *The Turkish Online Journal of Educational Technology*, 11(4), 295-306.
- Andrich, D. (1988a). *Rasch models for measurement*. Newbury Park, CA: Sage.
- Andrich, D. (1988b). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, 1(4), 363-378.
- Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath & S. H. Lovibond (Eds.), *Mathematical and theoretical systems* (pp. 7-16). North Holland: Elsevier Science.
- Andrich, D., Sheridan, B. E., & Luo, G. (2010). *RUMM2030: A windows-based item analysis program employing Rasch unidimensional measurement models*. Perth, Western Australia: RUMM Laboratory.
- Andrich, D., & van-Schoubroeck, L. (1989). The general health questionnaire: A psychometric analysis using latent trait theory. *Psychological Medicine*, 19, 469-485.

- Answer.Com. (2010). Aceh. Retrieved from <http://www.answers.com/topic/aceh>.
- Arshad, M., Ali, M. & Naseem, C. (2013). Does gender difference matter in learning English in a culture? *Interdisciplinary Journal of Contemporary Research in Business*, 5(3), 409-419.
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189-208.
- Baetens-Beardsmore, H. B. (2003). Who's afraid of bilingualism? In J.-M. Dewaele, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles*. (pp. 10-27). Clevedon, U. K.: Multilingual Matters.
- Baker, C. (1996). *Foundations of bilingual education and bilingualism* (2nd ed.). Clevedon, UK: Multilingual Matters.
- Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Bristol: UK: Multilingual Matters.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Bristol: Multilingual Matters.
- Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, 36(2), 1129-1147.
- BBC. (2005). Indonesia. Retrieved from http://news.bbc.co.uk/2/shared/spl/hi/world/04/asia_quake/quake_maps/html/1.stm.
- Bergman, M. M. (2008). *Advances in mixed-methods research: Theories and applications*. Thousand Oaks, C A: Sage.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13(2), 116-137.
- Bialystok, E. (2010). Global-local and Trail-Making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology*, 46(10), 93-105.
- Bialystok, E. (2011). Reshaping the mind: The benefit of bilingualism. *Journal of Experimental Psychology*, 65(4), 229-235.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122(1), 67-73. doi: 10.1016/j.cognition.2011.08.003
- Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptom of dementia. *Neuropsychologia*, 45, 459-464.

- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.
- Bialystok, E., Craik, F. I. M., & Ruocco, A. (2006). Dual-modality monitoring in a classification task: The effects of bilingualism and ageing. *The Quarterly Journal of Experimental Psychology, 59*(11), 1968-1983.
- Bialystok, E., & Herman, J. (1999). Does bilingualism matter for early literacy. *Bilingualism: Language and Cognition, 2*(1), 35-44.
- Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages bilingual children in two cultures. *Cognition, 112*, 494-500.
- Biedron, A., & Szczepaniak, A. (2012). Working memory and short-term memory abilities in accomplished multilinguals. *The Modern Language Journal, 96*, 290-306.
- Bolukbas, F., Keskin, F., & Polat, M. (2011). The effectiveness of cooperative learning on the reading comprehension skills in Turkish as a foreign language. *TOJET: The Turkish Online Journal of Educational Technology, 10*(4), 330-335.
- Brisk, M. E. (2005). Bilingual education. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97-113.
- Butler, Y. G. (2013). Bilingualism/multilingualism and second language acquisition. In T. K. Bathia & W. C. Ritchie (Eds.), *Handbook of bilingualism and multilingualism* (2nd ed., pp. 109-136). Somerset, NJ: Wiley.
- Buyuktaskapu, S. (2012). Effect of family supported pre-reading training program given to children in preschool education period on reading success in primary school. *Educational Sciences: Theory & Practice, 12*(1), 309-316.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282-298.
- Cazden, C. B., & Snow, C. E. (Eds.). (1990). *English plus: Issues in bilingual education*. London: Sage.
- Chapman, E. S. (2006). True score. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 1, pp. 421-423). Thousand Oaks, CA: Sage.
- Chodidjah, I. (2009). *Structured and monitored teacher development: The wind of change*. Paper presented at the the access English EBE symposium, Jakarta.

- Chow, B. W.-Y., McBride, C., Cheung, H., & Chow, C. S.-L. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology, 44*, 233-244.
- Clark, V. P., & Creswell, V. L. P. (2008). *The mixed methods reader*. Thousand Oaks, C A: Sage.
- Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. B. (2006). Prevalence of mixed methods sampling designs in social science research. *Evaluation and Research in Education, 19*(2), 83-101.
- Colome, A. (2001). Lexical activation in bilinguals' speech production: Language specific or language-independent? *Journal of Memory and Language, 45*(4), 721-736.
- Colzato, L. S., Bajo, M. T., Wildenberg, W. v. d., Paolieri, D., Nieuwenhuis, S., Heij, W. L., et al. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Recognition, 34*(2), 302-312.
- Conway, A. R. A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variation in working memory*. New York: Oxford University Press.
- Costa, A., Hernandez, M., & Sebastian-Galles, N. (2008). Bilingualism and conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59-86.
- Creswell, V. L. P., & Clark, V. P. (2006). *Designing and conducting mixed methods research*. Thousand Oaks, C A: Sage.
- Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish-English bilingual. *Applied Linguistics, 20*, 1-20.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
- Cummins, J. (2003). Bilingual education: Basic principles. In J.-M. Daelewe, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles. Festschrift in honour of Hugo Baetens Beardsmore* (pp. 56-66). Clevedon, England: Multilingual Matters.
- Davies, R. (2005, October 9). Rising to the challenge of bilingual education, online, *The Jakarta Post*. Retrieved from <http://www.thejakartapost.com/news/2005/07/16/bilingual-education-often-misunderstood-always-complicated.html>

- Depdiknas, N. E. D. (2007). *Pedoman penjaminan mutu sekolah/madrasah bertaraf internasional jenjang pendidikan dasar dan menengah*. Jakarta: Direktorat Tenaga Kependidikan, Direktorat Jenderal Peningkatan Mutu Pendidikan dan Tenaga Kependidikan, Departemen Pendidikan Nasional.
- Denscombe, M. (2010). *The good research guide: for a small-scale social research project* (4th ed.). Berkshire, UK: Open University Press.
- Diamond, A. (2002). Normal development of pre-frontal from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Strauss & R. Knight (Eds.), *Principle of frontal lobe function* (pp. 466-503). New York: Oxford.
- Dicker, S. J. (Ed.). (2003). *Languages in America: A pluralistic view* (2nd ed.). Clevedon: Multilingual Matters.
- Dixon, L. Q., Zhao, J., Shin, J.-Y., Wu, S., Su, J.-H., Burgess-Brigham, R., et al. (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, 82(5), 5-60.
- Education, G. (2010). Indian ocean tsunami. Retrieved from <http://www.globaleducation.edna.edu.au/globaled/go/pid/2258>
- Edwards, J. R. (2010). *Minority languages and group identity: Cases and categories*. Amsterdam: John Benjamins.
- Elvyanti, S. (n.d). Indonesian curriculum history. Retrieved from <http://file.upi.edu/Direktori/E%20-%20FPTK/JUR.%20PEND.%20TEKNIK%20ELEKTRO/197311222001122%20-%20SISCKA%20ELVYANTI/sejarah%20kurikulum1.pdf>
- Exchange., S. (n.d). The education system in Indonesia. Retrieved from <http://www.spainexchange.com/guide/ID-education.htm>
- Ferguson, C. A., Houghton, C., & Wells, M. H. (1977). Bilingual education: An international prespective. In B. Spolsky & R. Cooper (Eds.), *Frontier of bilingual education*. Rowley, MA: Newbury House.
- Forum, A. R. (2006). Mendiknas resmikan "Fatih Bilingual School". Retrieved from <http://www.acehrecoveryforum.org/id/index.php?action=PSCM&no=2437>
- Foursha-Stevenson, C., & Nicoladis, E. (2011). Early emergence of syntactic awareness and cross-linguistic influence in bilingual children' judgments. *International Journal of Bilingualism*, 15(4), 521-534.
- Foy, J. G., & Mann, V. A. (2013). Bilingual children show advantages in nonverbal auditory executive function task. *International Journal of Bilingualism*, 0(0), 1-13.

- Francis, N. (1999). Bilingualism, writing, and metalinguistic awareness: Oral-literate interaction between first and second languages. *Applied Psycholinguistics*, 20, 533-561.
- Freeman, R. (2007). Reviewing the research on language education programs. In O. Garcia & C. Baker (Eds.), *Bilingual education: An introductory reader*. Clevedon: Multilingual Matters (pp. 3-18).
- Freedman, S. E., & Burlow, J. (2011). Using whole-word production measures to determine the influence of phonotactic probability and neighborhood density on bilingual speech production. *International Journal of Bilingualism*, 16(4), 369-387.
- Gebhard, J. G. (2012). International students' adjustment problems and behaviors. *Journal of International Students*, 2(2), 184-193.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Rowley, M A: Newbury House.
- Ghonsooly, B., & Shawqi, S. (2012). The effects of foreign language learning on creativity. *English Language Teaching*, 5(4), 161-167.
- Goldstein, B. A., & Bunta, F. (2011). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, 16(4), 388-401.
- Government, A. (2000). Peraturan Daerah Provinsi Daerah Istimewa Aceh Nomor 4 Tahun 2000. Retrieved from http://www.acehprov.go.id/images/stories/file/Perda/perda_prov_nad_no_04_tahun_2000.pdf
- Grech, H., & McLeod, S. (2012). Multilingual speech and language and disorder. In D. Battle (Ed.), *Communication disorders and development in multicultural population* (4th ed., pp. 120-147). St. Louis, MO: Elsevier.
- Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, 28, 41-50.
- Greene, J. C., Caracelli, V. J., & Graham, W. E. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Holland, P. W., & Wainer, H. (1993) *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum
- Hornberger, N. H. (1991). Extending enrichment bilingual education: Revisiting typologies and redirecting policy. In O. Garcia (Ed.), *Bilingual education*

focusschrisft in honour of Joshua A. Fisherman (Vol. 1, pp. 125-234).
Philadelphia: John Benjamins.

- Hosseini, S. B. (2012). Asynchronous computer-mediated corrective feedback and the correct use of prepositions: Is it really effective? *Turkish online Journal of Distance Education, TOJDE*, 13(4), 95-111.
- Jafari, N. & Ansari, D. N. (2012). The effect of collaboration on Iranian EFL learners' writing ability. *International Studies*, 5(2), 125-131.
- Jallad, N.Y. & Bani Abdelrahman, A. A. (2008). The Effect of Multiple Intelligences Strategies on EFL Ninth Graders' Achievement in Reading Comprehension. *Indian Journal Of Applied Linguistics*, 34(1-2), 87.
- Jones, D. V., & Martin-Jones, M. (2004). Bilingual education and language revitalisation in Wales: Post achievement and current issues. In J. W. Tollefson & A. B. M. Tsui (Eds.), *Medium instruction policies: Which agenda? Whose Agenda?* Mahwah, NY: Erlbaum.
- Jorda, M. P. S. (2003). Metapragmatic awareness and pragmatic production of third language learners of English: A focus on request acts realizations. *International Journal of Bilingualism*, 7(1), 43-69.
- Judge, J. W. (2011). Use of language learning strategies by Spanish adults for business English. *International Journal of English Studies*, 12(1), 705-710.
- Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2011). The relationship between vocabulary and short-term memory measures in monolingual and bilingual speakers. *International Journal of Bilingualism*, 15(4), 408-425.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin Review*, 16(4), 705-710.
- Kave, G., Eyal, N., Shorek, A., & Cohen-Manfield, J. (2008). Multilingualism and cognitive state in the oldest old. *psychology and aging*, 23(1), 70-78.
- Keats, J. A. (1997). Classical test theory. In J. P. Keeves (Ed.), *Educational research, methodology and measurement* (2nd ed., pp. 713-719). Cambridge, U K: Pergamon Press.
- Kemp, C. (2007). Strategic processing in grammar learning: Do multilinguals use more strategies? *International Journal of Bilingualism*, 4, 241-261.
- Kharkhurin, A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilingual divergent thinking. *Bilingualism: Language and Cognition*, 11(2), 225-243.

- Kharkhurin, A. V. (2009). The role of bilingualism in creative performance on divergent thinking and invented alien creatures texts. *Bilingualism and Creativity*, 43(1), 59-71.
- Kieffer, M. J., & Lesaux, N. K. (2012). Development of morphological awareness and vocabulary knowledge in Spanish-speaking language minority learners: A parallel process latent growth curve model. *Applied Psycholinguistics*, 33(1), 23-54.
- Kirkpatrick, A. (2007). *World englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.
- Kirkpatrick, A. (2012). English in ASEAN: Implications for regional multilingualism. *Journal of Multilingual and Multicultural Development*, 33(4), 331-344.
- Kirkpatrick, A., & Sussex, R. (Eds.). (2012). *English as an international language in Asia: Implications for language education*. London, UK: Springer.
- Kluger, J. (2013). The power of the bilingual brain. *Time*, 182, 24-29.
- Knoch, U. (2010). Investigating the effectiveness of individualized feedback to rating behavior: A longitudinal study. *Language Testing*, 28(2), 179-200.
- Koh, K. H. (2008). *Using Rasch measurement to construct a diagnostic reading assessment battery*. Paper presented at the the third International Rasch conference, University of Western Australia, Perth.
- Kovacs, A. M., & Mehler, J. (2009). *Cognitive gains in 7-month old bilingual infants*. Paper presented at the the national academy of Sciences of the united states of America.
- Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, but not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism* 9(2), 119-135.
- Kroll, J. F., & deGroot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. D. Groot & J. F. Kroll (Eds.), *Tutorial in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Lauchlan, F., Parisi, M., & Fadda, R. (2012). Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a 'minority' language. *International Journal of Bilingualism*, 17(1), 43-56.
- Laurent, A., & Martinot, C. (2009). Bilingualism and phonological segmentation of speech: The case of English-French pre-schoolers. *Journal of Early Childhood Literacy*, 9, 29-49.

- Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review*, 63, 605-628.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543-560.
- Lin, A. M. Y., & Man, E. Y. F. (2009). *Bilingual education: Southeast Asia perspectives*. Hong Kong: Hong Kong University Press.
- Mackey, W. F. (1978). The importation of bilingual education models. In J. Alatis (Ed.), *Georgetown University Roundtable: International dimensions of education*. Washington DC: Georgetown University Press.
- Maher, J. (1997). Linguistic minorities and education in Japan. *Educational Review*, 49(2), 115-127.
- Maley, C., & Bond, T. G. (2011). Measuring cognitive development in early childhood learning environments. In R. F. Cavanaugh & R. F. Waugh (Eds.), *Application of Rasch measurement in learning environments research*. Rotterdam: Sense Publishers.
- Marashi, H., & Dadari, L. (2012). The impact of using task-based writing on EFL learners' writing on EFL learners' writing performance and creativity. *Theory and Practice in Language Studies*, 2(12), 2500-2507.
- Marian, V., Spivey, M., & Hirsh, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86(1), 70-82.
- Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4), 279-298.
- Masters, G. N. (1997). Partial credit model. In J. P. Keeves (Ed.), *Educational research, methodology and measurement* (2nd ed., pp. 857-863). Cambridge, U K: Pergamon Press.
- McMurray, A. J., Pace, R. W., & Scott, D. (2004). *Research: A common sense approach*. Victoria, W A: Thomson.
- Mekheimer, M. A. A. (2012). Assessing aptitude and attitude development in a translation skills course. *CALICO Journal*, 29(2), 321-340.
- Menanet. (n.d). National education. Retrieved from <http://home.menanet.net/~indoemb/indoemb/univ.htm>.
- Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, 23, 226-235.

- Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: multilingualism and multilingual education* (pp. 35-63). Clevedon, U K: Multilingual Matters.
- Metsämuuronen, J., Svedlin, R., & Ilic, J. (2012). Change in pupils' and students' attitude towards school as a function of age: A Finnish perspective. *Journal of Educational and Development Psychology*, 2(2), 134-151.
- Michel, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, N J: Lawrence Erlbaum.
- Michel, J. (1999). *Measurement in psychology: A critical history of methodological concept*. Cambridge, U K: Cambridge University Press.
- Miles, A., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, C A: Sage.
- Miyake, A., Friedman, N. P., Emerson, M. J., & Witzki, A. H. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A talent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Mohandas, R. (n.d). Overview of education for sustainable development (ESD) in Indonesia. Retrieved from www.ias.unu.edu
- Morales, J., Calvo, A., & Bialystok, A. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114, 187-202.
- Morse, M. M., & Niehaus, L. (2009). *Mixed method design: principles and procedures*. Walnut Creek, California: Left Coast.
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantages. *Developmental Science*, 10(6), 179-726.
- Murad Sani, A. & Zain, Z. (2011). Relating adolescents' second language reading attitudes, self-efficacy for reading, and reading ability in a non-supportive ESL setting. *Reading Matrix: An International Online Journal*, 11(3), 243-254.
- Nakanishi, T., & Ueda, A. (2011). Extensive reading and the effect of shadowing. *Reading in a Foreign Language*, 23(1), 1-16.
- Nasional, D. P. (2001). Kebijakan umum pendidikan dasar dan menengah. Retrieved from <http://sibi.or.id/download/Bing.doc>
- Nurdin, M. (2006). Banda Aceh overview after one year Tsunami. *Journal of Disaster Research*, 1(1), 116-117.
- O'Neill, S. (2008). A case study of learning in English as a foreign language (EFL) in Japan: High school students' English proficiency levels and fostering positive

cross-cultural attitudes. *International Journal of Pedagogy and Learning*, 4(5), 104-120.

Ovando, C. J., & Combs, M. C. (2012). *Bilingual and ESL Classrooms: Teaching in multicultural contexts* (5th ed.). New York: McGraw-Hill.

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the hospital anxiety and depression scale (HADS). *British Journal of Clinical Psychology*, 46, 1-18.

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effect of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, 108(3), 567-579.

Prior, A., & MacWhinney, B. (2009). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253-262.

Rahmani, M., & Sadeghi, K. (2011). Effects of note-taking training on reading comprehension and recall. *Reading Matrix: An International Online Journal*, 11(2), 116-129.

Ransdell, S., Barbier, M.-L., & Niit, T. (2006). Metacognitions about language skills and working memory among monolingual and bilingual college students: When does multilingualism matter? *The International Journal of Bilingual Education and Bilingualism*, 9, 728-741.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark's Paedagogiske Institut.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (expanded edition, originally published in 1960). Chicago, IL: Mesa Press.

Rasch, G. (2010). *Probabilistic models for some intelligence and attainment tests (50th anniversary ed.)*. Chicago, IL: Mesa Press.

Rauch, D. P., Naumann, J., & Jude, N. (2011). Metalinguistic awareness effects of full biliteracy on third-language reading proficiency in Turkish-German bilinguals. *International Journal of Bilingualism*, 16(4), 402-418.

Revianur, A. (2013a, 8 January). RSBI dan SBI jadi sekolah biasa, online, *Kompas.com*. Retrieved from http://edukasi.kompas.com/read/2013/01/08/18233422/RSBI.dan.SBI.Jadi.Sekolah.Biasa?utm_source=edukasi&utm_medium=cpc&utm_campaign=artbox

Revianur, A. (2013b, 8 January). Ini alasan MK batalkan status RSBI/SBI, online, *Kompas.com*. Retrieved from <http://edukasi.kompas.com/read/2013/01/08/18431250/Ini.Alasan.MK.Batalkan.Status.RSBISBI#komentar>

- Rispen, J. E., McBride-Chang, C., & Reitsma, P. (2008). Morphological awareness and early and advanced word recognition and spelling in Dutch. *Reading and Writing, 21*, 587-607.
- Rocco, T. S., Bliss, L. A., Gallagher, S., Perez-Prado, A., Alacaci, C., Dwyer, E. S., et al. (2003). The pragmatic and dialectical lenses: Two views of mixed methods use in education. In A. Tashakkori & C. Teddlie (Eds.), *The handbook of mixed methods in the social and behavioural sciences*. Thousand Oaks, CA: Sage.
- Rodriquez-Fornells, A., Balaguer, R. D., & Munte, T. F. (2006). Executive control in bilingual language processing. *Language Learning, 56*(1), 133-190.
- Rodriquez-Fornells, A., Rotte, M., Heinze, H. J., Nosselt, T., & Munte, T. F. (2002). Brain potential and functioning MRI evidence for how to handle two languages with one brain. *Nature, 415*, 1026-1029.
- Rossell, C. H., & Baker, K. (1996). The effectiveness of bilingual education. *Resources in the Teaching of English, 30*(1), 7-74.
- Roy, S., & Galiev, A. (2011). Discourses on bilingualism in Canadian French immersion programs *Canadian Modern Language Review, 67*(3), 351-376.
- RUMM 2030 Manual*. (2009). Perth, Western Australia: RUMM Laboratory.
- Sadeghi, K. & Sharifi, F. (2013). The effect of post-teaching activity type on vocabulary learning of elementary EFL learners (2013). *English Language Teaching, 6*(11), 65-76.
- Sahin, A. (2011). Effects of linear texts in page scrolling and page-by-page reading forms on reading comprehension introduction. *TOJET: The Turkish Online Journal of Educational Technology, 10*(3), 94-100.
- Shih, S. (2012). *Bilingualism in school and society: Language, identity, and policy*. New York: Taylor and Francis.
- Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement *Structural Equation Modelling* (Vol. 3, pp. 25-40).
- Soleimani, H., Mainnzadeh, A., Kassaian, Z., & Ketabi, S. (2012). The effect of instruction based on multiple intelligence theory on the attitude and learning of general English. *English Language Teaching, 5*(9), 45-53.
- Sumiya, H., & Healy, A. (2004). Phonology in the bilingual stroop effect. *Memory and Cognition, 32*, 752-758.
- Swanson, H. L., Orosco, M. J., Lussier, C. M., Gerber, M. M., & Guzman-Orth, D. A. (2011). The influence of working memory and phonological processing on English language learner children's bilingual reading and language acquisition. *Journal of Educational Psychology, 103*(4), 838-859.

- Syahdan, S. (2009). Developing characteristics of excellent teachers towards international standardised school. Retrieved from <http://pps.unnes.ac.id/pps1/files/paper%20seminar%20pbi%202009/Syahdan.doc>
- Tabatabaei, O., & Assefi, F. (2012). The effect of portfolio assessment technique on writing performance of EFL learners. *English Language Teaching*, 5(5), 138-147.
- Tajeddin, Z., & Daraee, D. (2013). Vocabulary acquisition through written input: Effects of form-focused, message oriented, and comprehension tasks. *Teaching English as a Second or Foreign Language*, 16(4), 1-19.
- Treccani, B., Argyri, E., Sorace, A., & Salla, S. D. (2009). Spatial negative priming in Bilingualism. *psychonomic Bulletin & Review*, 16(2), 320-327.
- Tubagus, A. (2005). Foreign teachers not qualified. *Asia Africa Intelligence Wire*. Retrieved from <http://www.accessmylibrary.com/article-1G1-136777930/foreign-teachers-not-qualified.html>
- Vorstman, E. L. P., Swart, H. d., Ceginskas, V., & Bergh, H. v. d. (2009). Language learning experience in school context and metacognitive awareness of multilingual children. *International Journal of Multilingualism*, 6(3), 258-280.
- Wang, C., Kim, D.-H., Bong, M., & Ahn, H. S. (2012). Examining measurement properties of an English self-efficacy scale for English language learners in Korea. *International of Educational Research*, 59, 24-34.
- Washington, C. (2010). Aceh. Retrieved from <http://courses.washington.edu/larescue/aceh6.htm>
- Waugh, R. F. (2007). Rasch measurement. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 820-825). Thousand Oaks, CA: Sage.
- Waugh, R. F. (Ed.). (2003). *On the forefront of educational psychology*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2005). *Frontiers in educational psychology*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2010a). *Specialized Rasch measures applied at the forefront education*. New York: Nova Science Publishers.
- Waugh, R. F. (Ed.). (2010b). *Applications of Rasch measurement in education*. New York: Nova Science Publishers.
- Waugh, R. F., Bowering, M., & Chayarathee, S. (2005). Creating scales to measure reading comprehension, and attitude and behaviour, for prathom 6 (grade 6) students taught ESL through a cooperative learning method in Thailand. In R. F.

Waugh (Ed.), *Frontiers in Education Psychology* (pp. 183-219). New York: Nova Science Publishers.

Waugh, R. F., Bowering, M. H., & Torok, S. (2005). Creating scales to measure reading comprehension, and attitude and behaviour, for prathom (grade) 7 students taught ESL through a genre-based method in Thailand. In R. F. Waugh (Ed.), *Frontiers in Education Psychology* (pp. 133-173). New York: Nova Science Publishers.

Waugh, R. F., & Chapman, E. (2005). An analysis of dimensionality using factor analysis (true score theory) and Rasch measurement: What is the difference? Which method is better. *Journal of Applied Measurement*, 6(1), 80-99.

Wei, L. (2000). *The bilingualism reader*. London: Routledge.

Wei, L. (2010). BAMFLA: Issues, methods and directions. *International Journal of Bilingualism*, 14(1), 3-9.

Wijaya, H. P. S. (2008). The impacts of changes in education policy on ELT in Indonesia. Retrieved from [http://fportfolio.petra.ac.id/user_files/04-036/The%20Impacts%20\(full%20complete\).doc](http://fportfolio.petra.ac.id/user_files/04-036/The%20Impacts%20(full%20complete).doc)

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modelling*, 3(1), 3-24.

Wu, Y. J., & Thierry, G. (2010). Investigating bilingual processing: The neglected role of language context. *Frontiers in Language Sciences*, 1, 1-6.

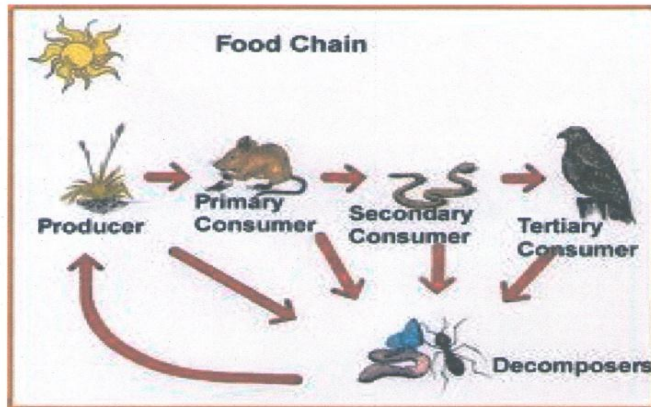
Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to difference in reading behaviour, text comprehension and noticing of language form? *Language Teaching Research*, 10(4), 419-434.

Yuwono, G. (2005). *English language teaching in decentralised Indonesia: Voices from the less privileged schools*. Paper presented at the AARE 2005 International Education Research Conference. <http://www.aare.edu.au/05pap/yuw05050pdf>

APPENDICES

Appendix A-English Reading Comprehension

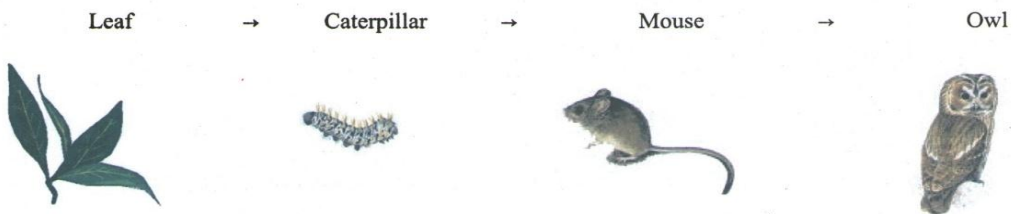
FOOD CHAINS



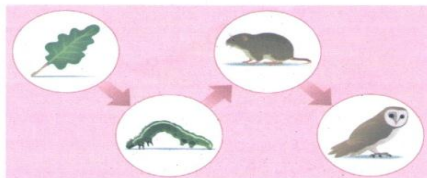
NEEDS FOR ENERGY

When you need energy in the morning before you come to school what do you do? Eat breakfast. Living things need food to give them energy. Plants get energy from sunlight. Caterpillars get their energy from plants. Mice get their energy from caterpillars. Owls get their energy from mice. Plants, caterpillars, mice and owls when die, give energy to decomposers. The plants and animals in a food chain depend on each other to live. What would happen if all the plants died? The caterpillars would die. All the animals that eat depend on the caterpillars would die too. So every part of the chain is important.

Look at this picture representing a food chain using arrows,



Or this picture showing links or chains.



The caterpillar eats a leaf

The mouse eats the caterpillar

The owl eats the mouse

These links between animals and plants are called **food chains**

The food chain consists of four main parts: the sunlight, producers, consumers, and decomposers.

SUNLIGHT AND PRODUCERS

The Sunlight. All food chains begin with the sunlight. Plants use the sunlight to grow.

Producers: Plants are called producers because they make their own food and feed herbivores.

CONSUMERS AND DECOMPOSERS

Consumers: Animals are called consumers because they eat other plants and animals. Consumers are called differently according to what they eat.

1. Some animals eat plants, they are called herbivores. Herbivores are also called **primary consumers**.
2. Some animals eat other animals, they are called carnivores. Carnivores are also called **secondary consumers**.
3. Some animals eat and harm other animals, they are called parasites.
4. Some animal eat dead animals, they are called scavengers.

Decomposers: When plants and animal died, some bacteria and fungi change them into gases such as carbon and nitrogen. These gases are then released back into the air, soil, or water.

HUMANS IN THE FOOD CHAINS

Human are called omnivores because they can eat plants and animals. Humans have superior brains so they often eat animals that are larger than themselves.

[Adopted from various sources]

DIRECTIONS:

1. Put **your name** on the answer sheet provided.
2. Read the passage carefully and cross (X) a, b, or c that suits your answer best.

QUESTIONS:

Needs for Energy

1. What will happen if there is no food?
 - a. No problem
 - b. Plants will die
 - c. Plants and animals will die
2. From where do plants get their energy?
 - a. Animals
 - b. Sunlight
 - c. Humans

3. A food chain is
 - a. A link between plants and plants
 - b. A link between plants and animals
 - c. A link between animals and animals

Sunlight and Producers

4. What do plants use to grow?
 - a. Animals
 - b. Sunlight
 - c. Humans
5. What do we call plants?
 - a. Decomposers
 - b. Consumers
 - c. Producers
6. Plants make
 - a. their own food
 - b. other animals' food
 - c. their own food and other animals' food

Consumers and decomposers

7. What are primary consumers?
 - a. Animals that eat plants
 - b. Animals that eat plants and other animals
 - c. Animals that eat other animals
8. What is an example of a carnivore?
 - a. Cow
 - b. Goat
 - c. Lion
9. Decomposers are
 - a. Goats that eat grass
 - b. Bacteria and fungi that eat dead things
 - c. Cats that eat fish

Humans in the Food Chains

10. What do humans eat?
 - a. Plants
 - b. Animals
 - c. Plants and animals
11. What do we call humans?
 - a. Omnivores
 - b. Herbivores
 - c. Carnivores

12. Is Food Chain important? Why? Or Why not? Write your answer below.

Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.

Thank you for your help. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix B-English Reading Comprehension's Ordered Scoring Scheme

Reading Comprehension Item		Least correct	Partly correct or partly incorrect	Most correct
		Score 1	Score 2	Score 3
Item difficulty				
Need for Energy				
1	What will happen if there is no food?	a	b	c
2	A food chain is	a	c	b
3	Where do plants get energy from?	c	a	b
Parts of the Food Chain				
4	What are primary consumers?	c	b	a
5	Primary consumers are also called	a	b	c
6	What is an example of herbivores?	c	a	b
Secondary Consumers in the Food Chain				
7	What are carnivores?	a	c	b
8	Decomposers are	c	a	b
9	Which statement is <u>correct</u> ?	a	b	c
Humans in the Food Chain				
10	What do humans eat?	a	b	c
11	What do we call humans?	b	c	a
12	Is Food Chain important? Why?			

Appendix C-English Writing Test

Name:

Name of School:

Name of Class:

.....

.....

.....

WRITING

TOPIC 1. : *My Idul Fitri* Holiday

Direction: 1. Put your name, name of school, and name of class on the top right of this page.

2. Write some paragraphs (the more paragraphs the better) on '*My Idul Fitri* Holiday'

Time: 40 minutes

My Idul Fitri Holiday

5

10

15

TOPIC 2. : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ' My Family'.

Time: 40 minutes

My Family

5

10

15

20

Appendix D-English Writing Scoring Rubric

		TEXT WRITING	EXCELLENT (Score 4)	GOOD (Score 3)	ADEQUATE (Score 2)	POOR (Score 1)
Paragraph(s) Organisation						
Easy	1	Main Topic	Main topic is clearly stated in a complete opening sentence with correct sentence structure.	Main topic is stated in an opening sentence but sentence structure is not correct AND/OR lacking detail.	Main topic is not clearly stated but sentence structure is correct.	Main topic is not clearly stated. Sentence structure is not correct.
Hard	2	Concluding Sentence	Concluding sentence restates topic with correct sentence structure.	Concluding sentence restates topic but sentence structure is not correct AND/OR lacking detail.	Concluding sentence is not clearly stated but sentence structure is correct.	Concluding sentence is not clearly stated. Sentence structure is not correct.
Harder	3	Supporting Sentences	Text has at least 3 detailed supporting sentences AND all sentences are on topic.	Text has at least 3 supporting sentences *Only 2 sentences are on topic. AND/OR *Sentences lack details	Text has at least 3 supporting sentences *Only 1 sentence is on topic. *Sentences lack details.	Text has supporting details BUT none are on topic.
Text Conventions						
Easy	4	Spelling	There are 0-2 spelling errors.	There are 3-4 spelling errors.	There are 5-6 spelling errors.	There are more than 6 spelling errors.
Hard	5	Punctuation/ Capitalisation	Text has 0-2 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 3-4 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has 5-6 errors in punctuation/ capitalisation, and noun-verb agreement.	Text has more than 6 errors in punctuation/ capitalisation, and noun-verb agreement.
Harder	6	Grammar	Text has 0-2 errors in noun-verb agreement.	Text has 3-4 errors in noun-verb agreement.	Text has 5-6 errors in noun-verb agreement.	Text has more than 6 errors in noun-verb agreement.
Text Quality						
Easy	7	Readability	Text is neat and readable with 0-2 marked out words or other corrections.	Text is neat and readable with 3-4 marked out words or other corrections.	Text is not neat and readable with 5-6 marked out words or other corrections.	Text is not neat and unreadable with numbers of marked out words or other corrections.
Hard	8	Style (Sentence fluency, e.g. varied length,	Text shows sentence fluency.	Text shows reasonable sentence fluency.	Text shows minimal sentence fluency.	Text lacking in sentence fluency.

		good flow rhythm, and varied structure)				
Harder	9	Text Enjoyability	Text is enjoyable to read	Text is quite enjoyable to read	Text is satisfactory to read	Text is not enjoyable to read

Appendix E-Attitude/Behaviour Questionnaire

Name	
------	--



QUESTIONNAIRE:

Students' attitudes and behaviour questionnaire in relation to learning English.

Please put your name in the box on right top of this page.

Please rate the 18 statements according to the following response format and place and check (√) the column corresponding to your attitude (prior to studying), your behaviour (during study) on appropriate line opposite each statement:

- | | |
|----------------------|---------------------------|
| Most or all the time | put check (√) on column 3 |
| Some of the time | put check (√) on column 2 |
| Never or rarely | put check (√) on column 1 |

Instruction: Check (√) in the one column of the 3 scales that you feel most describes your feeling. Any response in the questionnaire will not affect your score in the course.

Subheading: Tasks for Listening Example

Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English			(√)		(√)	

Check (√) on the questionnaire item, if you believe, prior to learning that *you most or all the time like to pay attention to someone speaking English*, check (√) on column 3; if *you some of the time achieve this*, check (√) on column 2.

Now, answer the questions on the next page.

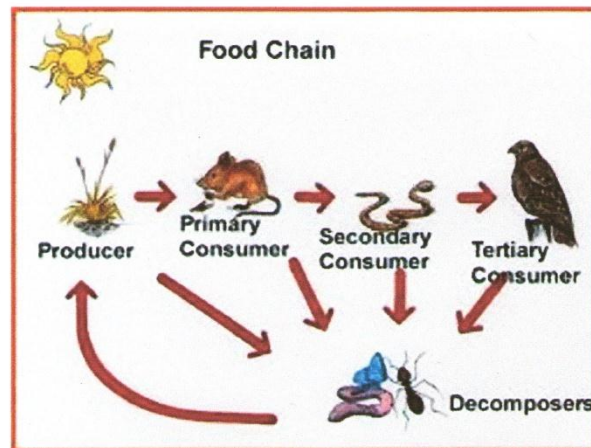
Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English						
3-4	I ask others to speak slowly or repeat words in English						
5-6	I listen to English songs						
	Tasks for Speaking						
7-8	I say new words several times in English						
9-10	I practice English with other students						
11-12	I start conversation in English with my friends						
	Tasks for Reading						
13-14	I guess the meaning of the English words in the text						
15-16	I read carefully words in English						
17-18	I can read English at home on my own						
	Tasks for Writing						
19-20	I look for similar words in my own language						
21-22	I divide English words into parts that I understand						
23-24	I write feelings in a diary in English						
	Student/student relationships						
25-26	I learn more when I study English in groups						
27-28	I can understand English better when I do activities with friends						
29-30	I like to lead my friends in doing English activities						
	Student/teacher relationships						
31-32	I learn a lot from my English teacher						
33-34	I like the way my teacher teaches English Reading Comprehension						
35-36	I like the way my teacher teaches English writing.						
	Common Views						
37-38	I like English because we use it in the classroom						
39-40	I like English because it helps me in higher study						
41-42	I like English because it helps me go abroad						

Is there anything else you would like to add about your experiences in learning English?

Thank you for your help.
It is appreciated. Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix F-Original English Reading Comprehension Text

FOOD CHAINS



Plants are important to our planet because they use photosynthesis to provide animal life with the energy they need to grow and to carry out life processes. When plants are eaten, this energy is transferred to another species. We call this a food chain.

There are three basic levels of a food chain.

Level 1—Plants are called producers because they have the ability to convert the Sun’s energy into chemicals that can be used to power life processes. Plants capture about 0.01 percent of the energy of the Sun’s rays every day and are the basis of the whole food chain.

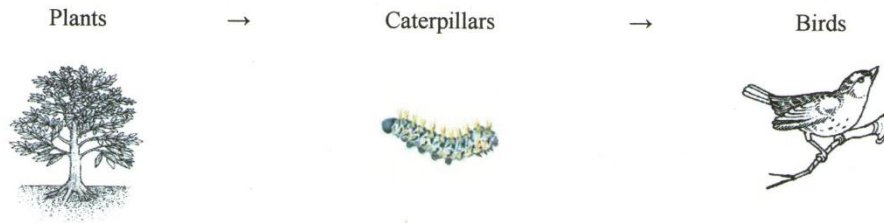
Level 2—Animals on the second level of a food chain are called primary consumers. We call them herbivores because they eat only plants for their energy. For example, deer eat shrub leaves, rabbits eat carrots and worms eat left litter. Primary consumers need to eat a lot of plants to meet their daily energy requirement. They eat approximately 10 per cent of the available plants and gain about 0.001 per cent of the Sun’s original energy.

Level 3—Animals on the third level are called secondary consumers. They eat primary consumers for their energy, for example, lions eat deer, foxes eat rabbits and birds eat worms. Secondary consumers are good predators, but there are fewer of them because their energy sources are limited. Secondary consumers usually eat meat and we call them carnivores (if secondary consumers eat mainly insects, we call them insectivores). Secondary consumers gain about 0.0001 per cent of the Sun’s original energy.

Further levels of the food chain are possible, but the longer a food chain is, the less energy left for the species at the top of the chain.

Representing food chains

We can represent food chains using arrows like this:



The arrow represents what species eat what.

That is:

Plants (eaten by) Caterpillar (eaten by) Birds

HUMANS AS CARNIVORES

Humans are at the top of their food chain because other animals do not naturally eat them. Humans are called omnivores because they can eat a variety of foods, such as plants and animals. Humans are not the largest of all species, but they have superior brains, they often eat animals that are larger than themselves. This is because they have used their intelligence to develop farming and hunting techniques.

[Source Adopted from: Walker, D (2006). *Adaptation and survival: Science Essential, Biology*. London: Evans]

DIRECTIONS:

1. Put **your name** on the **answer sheet** provided.
2. Read the passage carefully and cross (X) a, b, or c that suits your answer best.

QUESTIONS:

Plants in the Food Chain

1. To what does 'they' in line 1 refer?
 - a. Plants
 - b. Animals
 - c. Humans
2. A food chain is
 - a. a transfer of food from a species to another
 - b. a transfer of energy from a species to another
 - c. a transfer of chain from a species to another

Page | 3

Appendix G-Original Attitude/Behaviour Test

Questionnaire Item		This is what I wish to happen			This is what does happen		
		1	2	3	1	2	3
Item no.							
	Tasks for Listening						
1-2	I pay attention to someone speaking English						
3-4	I ask others to speak slowly or repeat words in English						
5-6	I listen to English songs						
	Tasks for Speaking						
7-8	I say new words several times in English						
9-10	I practice English with other students						
11-12	I start conversation in English with my friends						
	Tasks for Reading						
13-14	I guess the meaning of the English words in the text						
15-16	I read carefully words in English						
17-18	I can read English at home on my own						
	Tasks for Writing						
19-20	I look for similar words in my own language						
21-22	I divide English words into parts that I understand						
23-24	I write feelings in a diary in English						
	Student/student relationships						
25-26	I learn more when I study English in groups						
27-28	I can understand English better when I do activities with friends						
29-30	I like to lead my friends in doing English activities						
	Student/teacher relationships						
31-32	I learn a lot from my English teacher						
33-34	I like the way my teacher teaches English Reading Comprehension						
35-36	I like the way my teacher teaches English writing.						

Is there anything else you would like to add about your experiences in learning English?

Thank you for your help.

It is appreciated. Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix H-Request Letter to Head of Provincial Education of Aceh



September 27, 2010

Head
Dinas Pendidikan Aceh
in
Banda Aceh

Dear Sir,

With regard to activities for a PhD data collection which will be conducted by my PhD student, as the following details:

Name	:	Khairiah Syahabuddin
Student Number	:	10161350
Programme	:	Doctor of Philosophy (PhD)
Faculty	:	Education and Arts
University	:	Edith Cowan University, Mount Lawley Campus, Perth, Western Australia.

I would like to request that you can give permission to Khairiah Syahabuddin to collect her research data at Junior High School (Sekolah Menengah Pertama) in Banda Aceh for 1 (one) year, which meets a research procedure of Edith Cowan University, Australia.

Thank you. Your help and participation are appreciated.

A handwritten signature in cursive script, reading 'Russell Waugh', is positioned below the text of the letter.

Professor Russell Waugh

Principal Supervisor,
Edith Cowan University,
Mount Lawley Campus, 6050
Perth, Western Australia
Email: r.waugh@ecu.edu.au

Appendix I-Consent Letter to Headmaster



Dear Headmaster,

I am conducting a research project on student achievement in English Reading Comprehension; English writing; and Attitude and Behaviour.

The purpose of the project is to obtain information on your student achievement in English Reading Comprehension, English writing, and Attitude and Behaviour.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation of your students is voluntary and they can pull out at any time without prejudice. Your student participation has nothing to do with any formal or informal assessment in their school subjects.

Your student' name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of the students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

August 2010

I consent my students to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Annex I-Consent Letter to English Teacher



Dear Classroom Teacher,

I am conducting a research project on student achievement in English Reading Comprehension; English writing; and Attitude and Behaviour.

The purpose of the project is to obtain information on your student achievement in English Reading Comprehension, English writing, and Attitude and Behaviour.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation of your students is voluntary and they can pull out at any time without prejudice. Your student participation has nothing to do with any formal or informal assessment in their school subjects.

Your student' name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of the students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

August 2010

I consent my students to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix K-Consent Letter to student (English Reading Comprehension)



Dear Student,

I am conducting a research project on student achievement in English reading comprehension.

You are asked to complete the attached reading comprehension test. It contains 12 questions in multiple choices within 40 minutes.

The purpose of the test is to obtain information on your achievement in reading comprehension.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the test is given on the conditions mentioned above. Please return the completed test to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University (ECU), School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix L-Consent Letter to Student (English Writing)



Dear Student,

I am conducting a research project on student achievement in English writing.

You are asked to write some paragraphs (the more paragraphs is better) on two assigned topics. Each topic is allocated 35 minutes.

The purpose of the test is to obtain information on your achievement in writing.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@our.ecu.edu.au or khairiah_syahabuddin@yahoo.com

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the test is given on the conditions mentioned above. Please return the written paragraphs to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University, School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signed : _____

Printed Name : _____

Date : _____

Appendix M-Consent Letter to Student (Attitude/Behaviour Questionnaire)



Dear Student,

I am conducting a research project on student attitude and behaviour in learning English.

You are asked to complete the attached questionnaire. It contains 18 statements covering six aspects of your attitudes and behaviour in relation to learning English. These are tasks for listening, tasks for speaking, tasks for reading, tasks for writing, student/student relationships, and student/teacher relationships.

The purpose of the test is to obtain information on your attitude and behaviour in learning English.

It is expected that this research will benefit the schools, teachers, students, student advisors and researchers.

Participation is voluntary and you can pull out at any time without prejudice. Your participation has nothing to do with any formal or informal assessment in your school subjects.

Your name is required on the test sheets but individuals will remain anonymous. The research results will be published without the names of students or their schools, as they are only recorded on the computer program by numbers.

You can obtain a copy of the results or ask any questions about the study by contacting Khairiah Syahabuddin on +61403418644 (Australia) or +628126982942 (Indonesia) or ksyahabu@student.ecu.edu.au or khairiah_syahabuddin@yahoo.com

If you wish to speak with an independent person about the conduct of the project, please contact Ms Kim Gifkins, Research Ethics Officer, Edith Cowan University, on +61863042170, or research.ethics@ecu.edu.au

Thank you for your cooperation and participation. It is appreciated.

Your consent to complete the questionnaire is given on the conditions mentioned above. Please return the completed questionnaire to your English teacher or to Khairiah Syahabuddin.

Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)
PhD student of Edith Cowan University (ECU), School of Education,
2 Bradford Street, Mount Lawley, Western Australia 6050

October 2010

I consent to taking part in this research on the conditions stated.

Signature:

Name: _____

Date: _____

School name: _____

Appendix N-Student's Sample Answer of Reading Comprehension

ANSWER SHEET for Reading Comprehension 'FOOD CHAINS'

Name: _____

Name of School: _____

Name of Class: _____

Please answer here:

Read the passage carefully and cross (X) a, b, or c that suits your answer best.

Needs for Energy

1. a. b. c.
2. a. b. c.
3. a. b. c.

Sunlight and Producers

4. a. b. c.
5. a. b. c.
6. a. b. c.

Consumers and Decomposers

7. a. b. c.
8. a. b. c.
9. a. b. c.

Humans in the Food Chains

10. a. b. c.
11. a. b. c.

12. Is Food Chain important? Why? Or Why not? Write your answer below.

Yes, Because ~~living~~ living things need food to give their energy.

Plants get energy from sunlight. The plants and animals in a food chain depend on each other to live. So, every part of the chain is important.

Do you like English Reading Comprehension? Why? Or Why not? Write your answer below.

Yes, I like ^{english} reading comprehension.

Because:

1. I like english language.

2. English Reading Comprehension make me know about ~~comper~~ comprehension.

3. make me to be know much information

Thank you for your help. It is appreciated.
Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

**Appendix O-Student's Sample Answer English Writing
(Bilingual Student)**

Name: [Redacted]
Name of School: [Redacted]
Name of Class: [Redacted]

WRITING

TOPIC 1. : My Idul Fitri Holiday

Direction: 1. Put your name, name of school, and name of class on the top right of this page.
2. Write some paragraphs (the more paragraphs the better) on 'My Idul Fitri Holiday'

Time: 40 minutes

My Idul Fitri Holiday

5 Last year, I spent my Idul Fitri holiday in my village. I'm very happy in there. Because, I can play with my cousins and my friends. We're playing mud on the ricefield and we also made homework together. I have two cousins, they are very nice to me. We also celebrate my brother's birth on my village, and visited my grandfather's kuburan.

On first week of holiday, we was visit to my aunt house in Lhokseumawe. In there, we meet our family. With this time, I also to shop and buy many things in Suzuya store. Lhokseumawe is nice and good city. We can see fire on Pt. Arun, the fire is so beautiful.

10 Second week, we're back to Banda Aceh. Many memories to remember the holiday. Before that, I don't want to back to Banda Aceh. Because, I fell very happy in my village. I love my village.

15

TOPIC 2 : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ' My Family'.

Time: 40 minutes

My Family

- Hi... I want to tell you about my family. My name is [REDACTED].
I have a nicest family. My father's name is Asnawi Ismail. His job
is journalist on Harian Serambi Indonesia, and he is a lecture on
IAIN Ar-Raniry University. My father's favourite food is canai, same
5 to me. Because, canai is delicious food. Canai's from India. But,
we can buy that in "Canai Mamak Cafe". My father was born on
october, 10th, 1965.
- Now, I want to talk about my mother. My mother's name is Nur Azrah, sh.
She is a civil servant on KPU office. She has white skin and slim body.
10 Her hair is black and wavy. My mother's very like to make me many
food. like pizza, crocet, spaghetti, and many food's delicious. My mother
wear glasses, same to my father. Every day, my mother make me delicious
and healthy food.
- And one more, I have a little brother, his name is Hafiz Maulana,
15 He is 10 years old, he was born on January, 27th, 2001. He is
a smart boy, His school is elementary school number 67 Percontohan,
Banda Aceh. now, he is grade four. My little brother has pointed nose and
white skin. He's very handsome and nicest boy. He's very like otomotif,
like my uncle. In his bedroom, many toys of otomotif his collection.
- 20 Ok, guys, That's about my family. My family live at complex perumahan
Puklat Uke Kareng. If you want to Ulee Kareng, you can visit my house.
Thank you.

Appendix P-Student's Sample Answer English Writing (Monolingual Student)

Name: _____

Name of School: _____

Name of Class: _____

WRITING

TOPIC 1. : My Idul Fitri Holiday

Direction: 1. Put your name, name of school, and name of class on the top right of this page.
2. Write some paragraphs (the more paragraphs the better) on 'My Idul Fitri Holiday'

Time: 40 minutes

My Idul Fitri Holiday

My Name is _____, I am a student ~~the~~ ^{of} junior high school three, I am My Address is Ateuk Munjeng No. 5. my Father job is PNS (teacher), My Mother job is PNS (teacher).
at the feast of Idul Fitri, our family went to the mosque
5 after returning from the mosque we went to the house the young Sister to stay in touch. on our second holiday to return home. we went ^{up} ~~by~~ the car after I got ~~to~~ there, we greet shaking hands with the ~~our~~ grand mother, grand father. and other relatives.

TOPIC 2 : My Family

Direction: Write some paragraphs (the more paragraphs the better) on ' My Family'.

Time: 40 minutes

My Family

My Name is [REDACTED], My Mother ~~is~~ my name is
*****, my father name is ***** I have 2 young
Sister young sister 1. my name is NISA AUFIA, young sister 2 my name
is FATIN NAFISA, I am a student of junior-high school three
5 I am my adress is Ateul- Murijeng No. 5. is father Job.
PNS (teacher), My Mother job is (teacher matematika).
I happy the family. I am have unke my name is YAYAU
I am HAVE Aunt my Name is Xanti. I am is
very always swimming the pool - I Hobby Reading ~~book~~ Story
10 Book Malin Kundang, Laskar Pelangi, EVIL
old I am ~~the~~ ~~twelve~~, grand mother my name is NUSJANAH
My grand father Rusli M. ALI

15

20

Appendix Q-Student's Sample Answer Attitude/Behaviour Questionnaire (Bilingual Student)

Nama

Nama Sekolah

Nama Kelas

Pernyataan Kuesioner		Ini yang saya pikir seharusnya terjadi			Ini yang sebenarnya terjadi		
		1	2	3	1	2	3
Nomor Pernyataan							
Kegiatan kegiatan dalam Listening (mendengarkan)							
1-2	Saya mendengarkan dengan benar pada orang yang sedang berbicara Bahasa Inggris.			✓		✓	
3-4	Saya meminta orang lain untuk berbicara dengan lambat atau meminta untuk mengulangi kata kata ketika berbicara Bahasa Inggris.	✓				✓	
5-6	Saya mendengarkan lagu-lagu Bahasa Inggris.			✓			✓
Kegiatan-Kegiatan dalam Speaking (Berbicara)							
7-8	Saya mengucapkan kata kata baru Bahasa Inggris beberapa kali.			✓		✓	
9-10	Saya berlatih bicara Bahasa Inggris dengan siswa lain.					✓	
11-12	Saya memulai percakapan Bahasa Inggris dengan teman-teman saya.			✓		✓	
Kegiatan-Kegiatan dalam Reading Comprehension (Membaca dan Memahami)							
13-14	Saya menebak arti dari kata kata Bahasa Inggris dalam teks.		✓			✓	
15-16	Saya membaca kata kata Bahasa Inggris dengan seksama.			✓		✓	
17-18	Saya dapat membaca Bahasa Inggris sendiri di rumah.			✓		✓	
Kegiatan-Kegiatan dalam Writing (Menulis)							
19-20	Saya mencari kata kata Bahasa Inggris yang mirip dengan kata-kata Bahasa Indonesia ketika menulis Bahasa Inggris.	✓				✓	
21-22	Saya menulis kalimat Bahasa Inggris yang pendek pendek.	✓				✓	
23-24	Saya menulis perasaan saya ke dalam diari dalam Bahasa Inggris.			✓		✓	
Interaksi Siswa dengan Siswa							
25-26	Saya belajar banyak tentang Bahasa Inggris dalam kelompok belajar.			✓			✓
27-28	Saya dapat mengerti Bahasa Inggris dengan lebih baik ketika saya melakukan aktivitas dengan teman-teman saya.			✓		✓	
29-30	Saya suka memimpin teman-teman melakukan aktifitas Bahasa Inggris.			✓		✓	
Interaksi Siswa dengan Guru							
31-32	Saya belajar banyak tentang Bahasa Inggris dari guru Bahasa Inggris saya.			✓			✓
33-34	Saya suka cara guru Bahasa Inggris saya mengajar Membaca dan Memahami (Reading Comprehension).			✓			✓
35-36	Saya suka cara guru Bahasa Inggris saya mengajar Menulis (Writing).			✓			✓
Pandangan Umum							
37-38	Saya suka Bahasa Inggris karena kami menggunakannya di kelas.			✓			✓
39-40	Saya suka Bahasa Inggris karena dapat membantu saya ke sekolah lanjutan.			✓			✓
41-42	Saya suka Bahasa Inggris karena dapat membantu saya pergi ke luar negeri.			✓			✓

Ada hal-hal lain yang ingin kamu tambahkan tentang pengalamanmu dalam belajar Bahasa Inggris?

Orang tua saya sering membantu dalam belajar bahasa Inggris, misalnya Ibu saya sering mengajak saya berbicara bahasa Inggris, dan Ayah saya sering membelikan saya buku cerita berbahasa Inggris. Teman-teman saya di Facebook juga ada beberapa yg

Terima Kasih. Bantuanmu sangat dihargai. berasal dari luar negeri, jadi Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL) saya ber-chatting dengan mereka menggunakan bahasa Inggris.

Appendix R-Student's Sample Answer Attitude/Behaviour Questionnaire (Monolingual Student)

Nama [REDACTED]
 Nama Sekolah [REDACTED]
 Nama Kelas [REDACTED]

Pernyataan Kuesioner		Ini yang saya pikir seharusnya terjadi			Ini yang sebenarnya terjadi		
		1	2	3	1	2	3
Nomor Pernyataan							
Kegiatan kegiatan dalam Listening (mendengarkan)							
1-2	Saya mendengarkan dengan benar pada orang yang sedang berbicara Bahasa Inggris.		✓				✓
3-4	Saya meminta orang lain untuk berbicara dengan lambat atau meminta untuk mengulangi kata kata ketika berbicara Bahasa Inggris.		✓			✓	
5-6	Saya mendengarkan lagu-lagu Bahasa Inggris.			✓			✓
Kegiatan-Kegiatan dalam Speaking (Berbicara)							
7-8	Saya mengucapkan kata kata baru Bahasa Inggris beberapa kali.		✓			✓	
9-10	Saya berlatih bicara Bahasa Inggris dengan siswa lain.	✓				✓	
11-12	Saya memulai percakapan Bahasa Inggris dengan teman-teman saya.	✓			✓		
Kegiatan-Kegiatan dalam Reading Comprehension (Membaca dan Memahami)							
13-14	Saya menebak arti dari kata kata Bahasa Inggris dalam teks.		✓				✓
15-16	Saya membaca kata kata Bahasa Inggris dengan seksama.		✓			✓	
17-18	Saya dapat membaca Bahasa Inggris sendiri di rumah.			✓			✓
Kegiatan-Kegiatan dalam Writing (Menulis)							
19-20	Saya mencari kata kata Bahasa Inggris yang mirip dengan kata-kata Bahasa Indonesia ketika menulis Bahasa Inggris.	✓			✓		
21-22	Saya menulis kalimat Bahasa Inggris yang pendek pendek.		✓		✓		
23-24	Saya menulis perasaan saya ke dalam diari dalam Bahasa Inggris.		✓				✓
Interaksi Siswa dengan Siswa							
25-26	Saya belajar banyak tentang Bahasa Inggris dalam kelompok belajar.		✓			✓	
27-28	Saya dapat mengerti Bahasa Inggris dengan lebih baik ketika saya melakukan aktivitas dengan teman-teman saya.		✓		✓		
29-30	Saya suka memimpin teman-teman melakukan aktifitas Bahasa Inggris.			✓			✓
Interaksi Siswa dengan Guru							
31-32	Saya belajar banyak tentang Bahasa Inggris dari guru Bahasa Inggris saya.			✓			✓
33-34	Saya suka cara guru Bahasa Inggris saya mengajar Membaca dan Memahami (Reading Comprehension).		✓			✓	
35-36	Saya suka cara guru Bahasa Inggris saya mengajar Menulis (Writing).		✓			✓	
Pandangan Umum							
37-38	Saya suka Bahasa Inggris karena kami menggunakannya di kelas.		✓		✓		
39-40	Saya suka Bahasa Inggris karena dapat membantu saya ke sekolah lanjutan.			✓			✓
41-42	Saya suka Bahasa Inggris karena dapat membantu saya pergi ke luar negeri.			✓			✓

Ada hal-hal lain yang ingin kamu tambahkan tentang pengalamanmu dalam belajar Bahasa Inggris?

Untuk menambah wawasan saya, Untuk bisa membaca komik dan novel, mengerti arti dari lagu Inggris. Dan untuk membantu saya di masa depan.

Terima Kasih. Bantuanmu sangat dihargai.
 Khairiah Syahabuddin (S.Ag., M.HSc.ESL., M.TESOL)

Appendix S-Complete Occurrence of Student Written Comment

THEMES/CONCEPTS OF STUDENTS' WRITTEN COMMENTS

Themes/Concepts	Bilingual	Monolingual	Total
1. Tasks on Learning English as a Foreign Language			
Tasks for Listening			
I pay attention to someone speaking English	1	1	2
I ask others to speak slowly or repeat words in English	3	0	3
I like listening to English songs	25	19	44
Listening helps me to write	1	0	1
TOTAL			50
Tasks for Speaking			
I practice English with other students	9	1	10
It is important to be able to converse in English fluently	3	1	4
I like speaking in English	21	10	31
I like sing songs	3	1	4
TOTAL			49
Tasks for Reading			
I guess the meaning of the English words in the text	4	0	4
I like learning English through reading comprehension	16	8	24
I like reading books/e-books in English	5	0	5
TOTAL			33
Tasks for Writing			
I write feelings in a diary in English	1	1	2
I like to create a new sentence with a new vocabulary	1	0	1
I like to use dictionary for new words	1	1	2
I understand written text more than spoken	1	0	1
I like writing	2	2	4
TOTAL			10
2. Student-student relationships			
I learn more when I study English in groups	13	7	20
I ask friends to help me learning English	8	1	9
TOTAL			29
3. Student-teacher relationships			
I learn a lot from my English teacher	13	5	18
I like the way my teacher teaches English	57	41	98
I don't like the way my teacher teaches English	4	12	16
The way teacher teaches gives impact to students to love or hate English	9	3	12
TOTAL			144

4. Personal Views of Learning English as a Foreign Language			
Learning English is easy	15	5	20
Learning English is important	9	12	21
I like English because it is unique (it is pleased on ears but difficult to pronounce)	7	9	16
I like learning English	96	70	166
I like English because it is challenging	19	9	28
I like English because it is enjoyable/fun	74	50	124
I like English because number of good books are in English	1	0	1
I want to learn English with simple sentence	1	2	3
I am happy with my vocabulary	1	3	4
English is enjoyable if studied correctly	2	0	2
English is my favourite subject	1	0	1
Learning English needs time	2	1	3
I like making story in English	1	0	1
Other language helps understand English	1	1	2
I like to translate	0	3	3
I like storytelling	2	0	2
Dictionary helps me	7	3	10
TOTAL	408		
5. Common Views on Benefit of Learning English as a Foreign Language			
I like English because it helps me in higher study	76	59	135
I like English because it helps me go abroad/around the world	77	66	143
I like English because it helps me to study abroad	17	6	23
I like English because it helps me get scholarship for higher studies	1	0	1
I like English because it is an international language	46	41	87
I like English because it helps me in my daily activities	16	9	25
I like English as an additional language	3	7	10
I like English because it helps me to be enrolled in favourite schools	2	1	3
I like English because we use it in the classroom	6	0	6
I like English because it is a means to pursue my future dream jobs	48	27	75
TOTAL	508		
6. Students' Confidence and Achievement			
I am motivated to learn English	151	87	238
I take after-school English lesson to improve my English	18	7	25
I like English since I was a kid	4	0	4
I am lucky to understand English	1	0	1
I won English competition	11	7	18
TOTAL	286		

7. Learning English through Media			
I like watching TV programs in English for kids	1	0	1
I like watching TV programs/films in English	5	2	7
I like paying attention in found conversation in English movies	1	0	1
I like remembering sentences found in English movies	1	0	1
I like playing games/songs/story-book to improve my English	12	5	17
I like English because numbers of media/software/games/computers/video use English	7	2	9
I like learning English through English laboratory	1	0	1
TOTAL	37		
8. Family Support in Learning English as a Foreign Language			
My family (father, mother, uncles, aunts, siblings) encourage me to learn English well for future studies	37	6	43
I like to practice English with my older sister/brother	33	12	45
I like learning English to make my parents proud	0	7	7
TOTAL	95		
9. Obstacles in Learning English as a Foreign Language			
English is not easy	6	22	28
I do not understand vocab/meanings/grammar in English much	15	21	36
I do not like English because it is not easy to learn	31	41	72
I lack in English listening	2	3	5
I lack in English speaking	19	4	23
I lack in English writing	4	0	4
I lack in English reading comprehension	2	0	2
I do not like English, beside we do not need English in this country	4	0	4
I wish I knew English/good at it	3	5	8
I am not good in English	5	2	7
TOTAL	189		
10. Other Views			
The questionnaire gives me insights on my perceptions on English	2	0	2
The test given (by the researcher) is the first one considered well-prepared and presented	1	0	1
The test gave me lesson.	1	1	2
TOTAL	5		

1110 732 1846

**LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW
HKI**

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
Penulis : Khairiah Syahabuddin
Pengusul : Khairiah Syahabuddin
Status Pengusul : Pengusul tunggal
Identitas HKI : a. Penerbit : Kemenkumham RI
b. Tahun : 2017

Kategori : Konten HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
a. Kelengkapan unsur isi HKI (10%)	1	1
b. Ruang lingkup dan kedalaman pembahasan (30%)	3	3
c. Kecukupan dan kemutakhiran data/informasi dan metodologi (30%)	3	3
d. Kelengkapan unsur dan kualitas penerbit (30%)	3	3
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

a. Kelengkapan unsur isi HKI:

unsur isi lengkap

b. Ruang lingkup dan kedalaman pembahasan:

Ruang lingkup dan pembahasan mendalam

c. Kecukupan dan kemutakhiran data/informasi dan metodologi:

Data dan metodologi mutakhir

d. Kelengkapan unsur dan kualitas penerbit:

Berkualitas

Banda Aceh, 21 Februari 2022
Reviewer 1,


Prof. Dr. T. Zulfikar, M.Ed.

NIP. 197804302001121002

Guru Besar/Profesor Bidang English Language Research pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh

**LEMBAR PENILAIAN SEJAWAT SEBINDANG ATAU PEER REVIEW
HKI**

Judul HKI : Student English Achievement, Attitude and Behaviour in Bilingual and Monolingual Schools in Aceh, Indonesia
 Penulis : Khairiah Syahabuddin
 Pengusul : Khairiah Syahabuddin
 Status Pengusul : Pengusul tunggal
 Identitas HKI : a. Penerbit : Kemenkumham RI
 b. Tahun : 2017

Kategori : Konten HKI

Hasil Penilaian Peer Review:

Komponen yang Dinilai	Nilai Maksimum	Nilai Akhir yang Diperoleh
e. Kelengkapan unsur isi HKI (10%)	1	1
f. Ruang lingkup dan kedalaman pembahasan (30%)	3	3
g. Kecukupan dan kemutakhiran data/informasi dan metodologi (30%)	3	3
h. Kelengkapan unsur dan kualitas penerbit (30%)	3	3
Total (100%)	10	10
Nilai Pengusul (100%)=	10	10

Catatan Penilaian oleh Reviewer

- a. Kelengkapan unsur isi HKI:
sangat lengkap
-
- b. Ruang lingkup dan kedalaman pembahasan:
Lengkap dan mendalam
-
- c. Kecukupan dan kemutakhiran data/informasi dan metodologi:
Data/informasi dan metodologi mutakhir
-
- d. Kelengkapan unsur dan kualitas penerbit:
Lengkap dan berkualitas
-

Banda Aceh, 21 Februari 2022

Reviewer 2,

Dr. phil, Saiful Akmal, M.A.

NIP. 198203012008011006

Lektor Kepala Bidang Pendidikan Bahasa Inggris pada Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Ar-Raniry Banda Aceh