

**PEMBINAAN DAN PENGESAHAN INSTRUMEN
LITERASI PENTAKSIRAN PELAJAR IJAZAH
SARJANA MUDA PERGURUAN**

ZAKIAH BT. NOORDIN

*PERPUSTAKAAN
UNIVERSITI MALAYSIA SABAH*

**TESISINI DIKEMUKAKAN UNTUK
MEMENUHI SYARAT MEMPEROLEH IJAZAH
DOKTOR FALSAFAH**

**FAKULTI PSIKOLOGI DAN PENDIDIKAN
UNIVERSITI MALAYSIA SABAH
2016**

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL: **PEMBINAAN DAN PENGESAHAN INSTRUMEN LITERASI PENTAKSIRAN PELAJAR IJAZAH SARJANA MUDA PERGURUAN**

IJAZAH: **DOKTOR FALSAFAH (PENILAIAN DALAM PENDIDIKAN)**

Saya **ZAKIAH BT NOORDIN**, Sesi Pengajian **2011-2016**, mengaku membenarkan tesis Doktor Falsafah ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

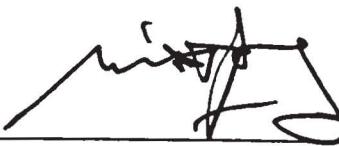
- SULIT** (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
- TERHAD** (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
- TIDAK TERHAD**



ZAKIAH BT NOORDIN
PT20119075

Disahkan oleh,
NURULAIN BINTI ISMAIL

• LIBRARIAN
UNIVERSITI MALAYSIA SABAH
(Tandatangan Pustakawan)



(Prof Dr. Vincent Pang
Penyelia

Tarikh: 20 Jun 2016



PENGAKUAN

Karya ini adalah hasil kerja saya sendiri kecuali nukilan, ringkasan dan rujukan yang setiap satunya dijelaskan sumber-sumbernya

10 Jun 2016



Zakiah bt Noordin
PT20119075



PENGESAHAN

NAMA : **ZAKIAH BT NOORDIN**

NO. MATRIK : **PT20119075**

TAJUK : **PEMBINAAN DAN PENGESAHAN INSTRUMEN LITERASI
PENTAKSIRAN PELAJAR IJAZAH SARJANA MUDA
PERGURUAN**

IJAZAH : **DOKTOR FALSAFAH (PENILAIAN DALAM PENDIDIKAN)**

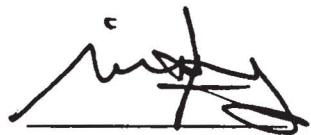
TARIKH VIVA : **14 MAC 2016**

DISAHKAN OLEH;

1. PENYELIA UTAMA

Prof Dr. Vincent Pang

Tandatangan



2. PENYELIA BERSAMA

Dr Mohd Khairuddin Abdullah @ Jerry



PENGHARGAAN

Alhamdulillah syukur kepada Allah SWT kerana dengan izin dan limpah kurniaNya tesis ini telah dapat disiapkan. Saya ingin merakamkan penghargaan dan ucapan terima kasih yang tidak terhingga kepada penyelia saya iaitu Prof. Dr Vincent Pang dan Dr Mohamad Khairuddin Abdullah yang telah membimbing dan memberi nasihat tanpa mengenal erti penat lelah.

Ucapan terima kasih juga diucapkan khususnya kepada Bahagian Tajaan Pendidikan, Institut Pendidikan Guru Malaysia (IPGM), rakan-rakan di IPG Kampus di seluruh Malaysia, pelajar-pelajar Program Ijazah Sarjana Muda Perguruan (PISMP) Ambilan Januari 2011 di IPG kampus seluruh Malaysia, rakan-rakan di IPG Kampus Gaya, Puan Juliana Osong, Dr. Akbariah Mohd Mahdzir, Dr. Connie Shin Ompok, Dr. Ibrahim Mohamed Zin, Dr. Zarina Ahmad, Dr. Lee Kean Wah, Dr. Jamil Ahmad, Prof Hairul Anwar Mastor, Dr. Othman Talib, Prof Ramayah dan Cik Norafidah Noralidin. Ucapan terima kasih kepada Puan Ling Mei Teng, Puan Yusnaili, Cik Siti Haswa Niza, Puan Siti Ajar Md Noh dan rakan-rakan seperjuangan yang tidak jemu-jemu mendoakan kejayaan saya.

Akhir sekali, berkat doa, kesabaran dan kasih sayang keluarga saya iaitu suami tersayang, Dr. Mohamad Nizam Nazarudin, anak-anak iaitu Nur Adli dan Nur Adila , kedua ibu bapa saya dan mertua.

ABSTRAK

Kajian ini bertujuan membina dan mengesahkan Instrumen Literasi Pentaksiran Institut Pendidikan Guru (ILP IPG) untuk mengukur literasi pentaksiran pelajar Ijazah Sarjana Muda Perguruan (PISMP). Berdasarkan analisis pengkaji terhadap *Standard for Teachers Competence in the Educational Assessment of Students (STCEAS)* dan analisis sukanan pelajaran kursus pengukuran dan penilaian di Institut Pendidikan Guru, lima dimensi literasi pentaksiran telah dibentuk iaitu Konsep Asas, Perancangan, Pembinaan, Pentadbiran, Penskoran dan Analisis; dan Interpretasi dan Tindakan Susulan. Proses pembinaan instrumen melibatkan tiga peringkat iaitu mereka bentuk; pembinaan dan pengesahan instrumen. Pembinaan ILP IPG juga berasaskan Model Pembinaan McIntire dan Miller. Kesemua item ILP IPG mempunyai kesahan yang baik kerana menunjukkan nilai *point-measure correlation* (PTMEA CORR) melebihi 0.3 dan nilai *fit* yang diterima. Nilai kebolehpercayaan item bagi kelima-lima dimensi ILP IPG mempunyai nilai yang melebihi 0.80. Indeks pengasingan item bagi kelima-lima dimensi adalah baik kerana melebihi nilai 2. Dapatkan kajian ini menunjukkan nilai kebolehpercayaan individu bagi kelima-lima dimensi adalah diterima iaitu melebihi nilai 0.6. Indeks pengasingan individu agak rendah kerana mempunyai nilai yang kurang dari nilai 2. Memandangkan wujudnya lompang-lompang dalam tafsiran konstruk yang diukur, maka indeks pengasingan yang agak rendah adalah dijangkakan. Analisis Kebezaan Fungsi Item ke atas jantina (*Gender Differential Item Functioning – GDIF*) mendapat 12 item menunjukkan GDIF signifikan dan ia perlu disemak semula bagi memastikan instrumen bersih dari sebarang item yang cenderung bias. Dalam kajian ini, item telah digubal dengan baik dan mensasari kebolehan responden yang diharapkan.

ABSTRACT

THE DEVELOPMENT AND VALIDATION OF ASSESSMENT LITERACY INSTRUMENT FOR BACHELOR OF TEACHING STUDENTS IN INSTITUT PENDIDIKAN GURU

The purpose of this research was to develop and validate Assessment Literacy Instrument for Teacher Education Institutes (ILP IPG) for use in the measurement of assessment literacy among Bachelor of Teaching students in Institut Pendidikan Guru. Based on the Standards for Teacher Competence in Educational Assessment of Students (STCEAS) and analysis of measurement and evaluation course syllabus, five dimensions of assessment literacy were developed namely Basic Concept, Planning, Designing, Administration, Scoring and Analysis and Interpretation and Follow-up. A three- stage process was utilised in the development of the instrument: (i) Instrument Design Stage, (ii) Item Writing Stage, and (iii) Validation Stage. The ILP IPG was designed based on the test development principles of McIntire and Miller. All items in ILP IPG showed a higher degree of validity based on PTMEA CORR values exceeding 0.3 and Outfit MNSQ and Infit MNSQ values between 0.7 and 1.3. The item reliability index for all five dimensions in ILP IPG had a value in excess of 0.80 with the item separation index is higher than 2. The findings of this study indicated the individual reliability index for these five dimensions to be between 0.65 to 0.78 with the separation index relatively low. This was expected in view of the existence of voids in the voids constructs measured interpretation. Gender Differential Item Functioning Analysis (GDIF) found 12 items showing the t value of DIF greater than 1.96. These items therefore should be reviewed to ensure an instrument free from any item bias. In this study, the items were developed properly and targeted the ability of respondents.

SENARAI KANDUNGAN

	Halaman
TAJUK	i
PENGAKUAN	ii
PENGESAHAN	iii
PENGHARGAAN	iv
ABSTRAK	v
ABSTRACT	vi
SENARAI KANDUNGAN	vii
SENARAI JADUAL	xv
SENARAI RAJAH	xvii
SENARAI SINGKATAN	xxii
SENARAI LAMPIRAN	xxiii
 BAB 1: PENGENALAN	 1
1.1 Pendahuluan	1
1.2 Latar Belakang Kajian	4
1.3 Pernyataan Masalah	6
1.4 Tujuan Kajian	10
1.5 Objektif Kajian	11
1.6 Soalan Kajian	11
1.7 Kepentingan Kajian	12
1.8 Batasan Kajian	14
1.9 Definisi Konsep Dan Operasi	14
1.9.1 Pembinaan	14
1.9.2 Pengesahan	15
1.9.3 Kesahan	15
1.9.4 Kebolehpercayaan	15
1.9.5 Literasi Pentaksiran	16
1.9.6 Instrument Literasi Pentaksiran	16
1.9.7 Dimensi Konsep Asas	16

1.9.8 Dimensi Perancangan	16
1.9.9 Dimensi Pembinaan	16
1.9.10 Dimensi Pentadbiran, Penskoran Dan Analisis	17
1.9.11 Dimensi Interpretasi Dan Tindakan Susulan	17
1.10 Rumusan	17
 BAB 2: KAJIAN LITERATUR	 18
2.1 Pengenalan	18
2.2 Pentaksiran	18
2.2.1 Literasi Pentaksiran	19
2.2.2 Model Kompetensi Pentaksiran	20
2.2.3 Kemahiran Menjalankan Pentaksiran Dalam Bilik Darjah	24
2.2.4 Konsep Asas Dan Terminology Dalam Pentaksiran	24
2.2.5 Kegunaan Pentaksiran	26
2.2.6 Merancang Dan Membina Kaedah Pentaksiran	26
2.2.7 Menginterpretasikan Hasil Pentaksiran	30
2.2.8 Mendeskripsi Hasil Pentaksiran	31
2.2.9 Menilai Menambahbaik Proses Pentaksiran	32
2.2.10 Penggredan Dan Tindakan Susulan	33
2.2.11 Etika Pentaksiran	33
2.3 Taksonomi Objektif Pengajaran	34
2.3.1 Taksonomi Objektif Kognitif	34
2.3.2 Taksonomi Bloom Dua Dimensi	35
2.4 System Pentaksiran Pendidikan Kebangsaan(Sppk)	38
2.4.1 Latar Belakang SPPK	38
2.4.2 Kaedah Pentaksiran Dalam SPPK	39
2.4.3 Pelaksanaan SPPK	43
2.5 Pelan Pembangunan Pendidikan Malaysia (PPPM)	45
2.6 Instrumen-Instrumen Literasi Pentaksiran	46
2.6.1 Instrumen Teacher Assessment Literacy Questionnaire (TALQ)	46
2.6.2 Instrumen Assessment Literacy Inventory (ALI)	47



2.6.3	Instrumen Classroom Assessment Literacy Inventory (CALI)	48
2.6.4	Instrumen Ujian Literasi Pentaksiran	49
2.7	Teori Ujian Klasik	53
2.7.1	Ralat Pengukuran	54
2.7.2	Kesahan	55
a.	Kesahan Kandungan	55
b.	Kesahan Kriteria	56
c.	Kesahan Konstruk	56
2.7.3	Kebolehpercayaan	57
a.	Uji-Uji Semula	58
b.	Kaedah Ujian Setara	58
c.	Ketekalan Dalaman	59
2.7.4	Indeks Kesukaran Item	61
2.7.5	Indeks Diskriminasi Item	62
2.7.6	Analisis Faktor	63
2.7.7	Kekangan Teori Ujian Klasik Dan Justifikasi Pengaplikasian Model Rasch	63
2.8	Teori Ujian Moden	65
2.8.1	Model Rasch	66
2.8.2	Kesahan Dalam Model Rasch	67
2.8.3	Polariti Item (Item Polarity)	68
2.8.4	Kesesuaian Item (Item Fit)	68
2.8.5	Peta Item- Individu (Item Map)	70
2.8.6	Menilai Ketidakpadanan (Misfit)	71
2.8.7	Kesesuaian Individu (Person Fit)	71
2.8.8	Kebolehpercayaan	71
2.8.9	Indeks Pengasingan (Separation Index)	73
2.8.10	Analisis Komponen Asas Bagi Kesamaan Residual Dimensi (Principal Component Analysis Of Residuals Unidimensionality)	74
2.8.11	Skala Pemeringkatan (Rating) Dan Keserasian Kategori	74

2.8.12	Analisis Kebezaan Fungsi Item (Differential Item Functioning - DIF)	75
2.9	Prosedur Pembinaan Instrumen	77
2.10	Model Pembinaan Ujian	79
2.10.1	Model Pembinaan Ujian Brown (1983)	80
2.10.2	Model Pembinaan Ujian Cohen Dan Swerdlik (2002)	80
2.10.3	Model Pembinaan Ujian Mcintire Dan Miller (2007)	83
2.11	Kerangka Konseptual Kajian	85
2.12	Rumusan	88
BAB 3: METODOLOGI		89
3.1	Pengenalan	89
3.2	Reka Bentuk Kajian	89
3.2.1	Peringkat I: Mereka Bentuk Instrumen	90
3.2.2	Peringkat II :Pembinaan Instrumen	91
3.2.3	Peringkat III: Pengesahan Instrumen	93
3.3	Populasi Dan Sampel Kajian	93
3.3.1	Pemilihan Sampel Dan Item Berdasarkan Model Rasch	96
3.4	Instrumen ILP IPG	98
3.4.1	Dimensi Konsep Asas	100
3.4.2	Dimensi Perancangan	101
3.4.3	Dimensi Pembinaan	101
3.4.4	Dimensi Pentadbiran, Penskoran Dan Analisis	101
3.4.5	Dimensi Interpretasi Dan Tindakan Susulan	101
3.4.6	Jadual Spesifikasi Ujian (JSU)	104
3.4.7	Format Tadbiran Ujian	108
3.4.8	Membina Item	108
3.4.9	Menyediakan Arahan Tadbiran	109
3.4.10	Arahan Kepada Pengawas Peperiksaan	109
3.4.11	Arahan Kepada Calon	109
3.4.12	Semakan Pakar	110
3.4.13	Kesahan Kandungan	111

3.5	Prosedur Pengumpulan Data	114
3.6	Penganalisisan Data	114
3.7	Rumusan	116
BAB 4: DAPATAN KAJIAN		117
4.1	Pengenalan	117
4.2	Dapatan ILP IPG Versi 1	117
4.2.1	Kebolehpercayaan Dan Indeks Pengasingan	118
4.2.2	Taburan Kesukaran Item Dan Kebolehan Responden	119
4.3	Dapatan ILP IPG Versi 1 (Dimensi Konsep Asas)	121
4.3.1	Kesahan (Polariti Item)	121
4.3.2	Kesahan (Kesesuaian Item)	122
4.3.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	124
4.3.4	Kebolehpercayaan Dan Indeks Pengasingan Item	126
4.3.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	126
4.3.6	Taburan Kesukaran Item Dan Kebolehan Responden	127
4.3.7	Unidimensi	129
4.4	Dapatan ILP IPG Versi 1 (Dimensi Perancangan)	130
4.4.1	Kesahan (Polariti Item)	130
4.4.2	Kesahan (Kesesuaian Item)	131
4.4.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	135
4.4.4	Kebolehpercayaan Dan Indeks Pengasingan Item	135
4.4.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	136
4.4.6	Taburan Kesukaran Item Dan Kebolehan Responden	137
4.4.7	Taburan Kesukaran Item Dan Kebolehan Responden	139
4.4.8	Unidimensi	141
4.5	Dapatan ILP IPG Versi 1 (Dimensi Pembinaan)	142
4.5.1	Kesahan (Polariti Item)	142
4.5.2	Kesahan (Kesesuaian Item)	143
4.5.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	148
4.5.4	Kebolehpercayaan Dan Indeks Pengasingan Item	148
4.5.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	149

4.5.6	Taburan Kesukaran Item Dan Kebolehan Responden	150
4.5.7	Unidimensi	152
4.6	Dapatan ILP IPG Versi 1 (Dimensi Pentadbiran, Penskoran Dan Analisis)	153
4.6.1	Kesahan (Polariti Item)	153
4.6.2	Kesahan (Kesesuaian Item)	154
4.6.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	159
4.6.4	Kebolehpercayaan Dan Indeks Pengasingan Item	160
4.6.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	161
4.6.6	Sasaran Alat Ukur (Test Targeting)	161
4.6.7	Taburan Kesukaran Item Dan Kebolehan Responden	162
4.6.8	Unidimensi	164
4.7	Dapatan ILP IPG Versi 1 (Dimensi Interpretasi Dan Tindakan Susulan)	166
4.7.1	Kesahan (Polariti Item)	166
4.7.2	Kesahan (Kesesuaian Item)	166
4.7.3	Taburan Kesukaran Item Dan Kebolehan Responden	169
4.7.4	Kebolehpercayaan Dan Indeks Pengasingan Individu	171
4.7.5	Kebolehpercayaan Dan Indeks Pengasingan Item	172
4.7.6	Ketepatan Anggaran Dan Sasaran Alat Ukur	172
4.7.7	Unidimensi	173
4.8	Dapatan Kajian ILP IPG Versi 2	174
4.8.1	Profil Responden	174
4.8.2	Kebolehpercayaan Dan Indeks Pengasingan Individu	175
4.8.3	Kebolehpercayaan Dan Indeks Pengasingan Item	176
4.8.4	Taburan Kesukaran Item Dan Kebolehan Responden	177
4.9	Dapatan ILP IPG Versi 2 (Dimensi Konsep Asas)	179
4.9.1	Kesahan (Polariti Item)	179
4.9.2	Kesahan (Kesesuaian Item)	180
4.9.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	180
4.9.4	Kebolehpercayaan Dan Indeks Pengasingan Item	181
4.9.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	182

4.9.6	Taburan Kesukaran Item Dan Kebolehan Responden	183
4.9.7	Unidimensi	185
4.10	Dapatan ILP IPG Versi 2 (Dimensi Perancangan)	186
4.10.1	Kesahan (Polariti Item)	186
4.10.2	Kesahan (Kesesuaian Item)	187
4.10.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	188
4.10.4	Kebolehpercayaan Dan Indeks Pengasingan Item	189
4.10.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	189
4.10.6	Taburan Kesukaran Item Dan Kebolehan Responden	190
4.10.7	Unidimensi	193
4.11	Dapatan ILP IPG Versi 2 (Dimensi Pembinaan)	194
4.11.1	Kesahan (Polariti Item)	194
4.11.2	Kesahan (Kesesuaian Item)	196
4.11.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	196
4.11.4	Kebolehpercayaan Dan Indeks Pengasingan Item	197
4.11.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	197
4.11.6	Taburan Kesukaran Item Dan Kebolehan Responden	198
4.11.7	Unidimensi	200
4.12	Dapatan ILP IPG Versi 2 (Dimensi Pentadbiran, Penskoran Dan Analisis)	201
4.12.1	Kesahan (Polariti Item)	201
4.12.2	Kesahan (Kesesuaian Item)	202
4.12.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	203
4.12.4	Kebolehpercayaan Dan Indeks Pengasingan Item	204
4.12.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	205
4.12.6	Taburan Kesukaran Item Dan Kebolehan Responden	205
4.12.7	Unidimensi	208
4.13	Dapatan ILP IPG Versi 2 (Dimensi Interpretasi Dan Tindakan Susulan)	209
4.13.1	Kesahan (Polariti Item)	209
4.13.2	Kesahan (Kesesuaian Item)	210
4.13.3	Kebolehpercayaan Dan Indeks Pengasingan Individu	211

4.13.4	Kebolehpercayaan Dan Indeks Pengasingan Item	212
4.13.5	Ketepatan Anggaran Dan Sasaran Alat Ukur	213
4.13.6	Taburan Kesukaran Item Dan Kebolehan Responden	213
4.13.7	Unidimensi	216
4.13.8	Perbandingan Dapatan Kajian ILP IPG Versi 1 Dengan Versi 2	217
4.14	Kesahan Respon Responden	220
4.15	Analisis Kebezaan Fungsi Item (Differential Item Functioning - DIF)	220
4.16	Rumusan	225
BAB 5 : PERBINCANGAN, CADANGAN DAN KESIMPULAN		226
5.1	Pengenalan	226
5.2	Perbincangan Dapatan Kajian	227
5.2.1	Dapatan Proses Pembentukan Konstruk Dan Penulisan Item	227
5.2.2	Adakah Item-Item Dalam Dimensi ILP IPG Menunjukkan Bukti Kesahan?	228
5.2.3	Adakah Item-Item Dalam Dimensi ILP IPG Menunjukkan Kebolehpercayaan?	229
5.2.4	Adakah Dikesan DIF Berdasarkan Jantina?	231
5.3	Implikasi Kajian	231
5.3.1	Implikasi Teoritikal	232
5.3.2	Implikasi Praktikal	234
5.4	Cadangan Kajian Lanjutan	234
5.5	Rumusan	235
RUJUKAN		236
LAMPIRAN		248

SENARAI JADUAL

	Halaman
Jadual 1.1 : Analisis Instrumen Pentaksiran Dalam dan Luar Negara	9
Jadual 2.1 : Perincian Kandungan STCEAS	22
Jadual 2.2 : Aras-aras Domain Kognitif	35
Jadual 2.3 : Struktur Dimensi Pengetahuan	36
Jadual 2.4 : Struktur Dimensi Proses Kognitif	37
Jadual 2.5 : Pelan Pelaksanaan Transformasi	44
Jadual 2.6 : Rumusan Analisis Instrumen Literasi Pentaksiran	52
Jadual 2.7 : Interpretasi MNSQ	69
Jadual 2.8 : Huraian julat z-std	70
Jadual 2.9 : Hubungan antara Separation dan Kebolehpercayaan	73
Jadual 2.10 : Perbandingan Aktiviti Antara Tiga Model Pembinaan Ujian	84
Jadual 3.1 : Cadangan Pemilihan Sampel	98
Jadual 3.2 : Dimensi Literasi Pentaksiran	99
Jadual 3.3 : Huraian Dimensi dalam ILP IPG	103
Jadual 3.4 : Perwakilan Item IP IPG	104
Jadual 3.5 : Jadual Spesifikasi Ujian	105
Jadual 3.6 : Jadual Spesifikasi Ujian Berdasarkan STCEAS	106
Jadual 3.7 : Maklumat Latar Belakang Pakar	110
Jadual 3.8 : Penilaian Keseluruhan Pakar	112
Jadual 3.9 : Senarai Item yang Diubahsuai Selepas Semakan Pakar	113
Jadual 3.10 : Kaedah Analisis Data	115
Jadual 4.1 : Senarai Item Sebelum dan Selepas Diubahsuai (Dimensi Konsep Asas)	124
Jadual 4.2 : Senarai Item Sebelum dan Selepas Diubahsuai (Dimensi Perancangan)	134
Jadual 4.3 : Senarai Item Sebelum dan Selepas Diubahsuai (Dimensi Pembinaan)	147
Jadual 4.4 : Senarai Item Sebelum dan Selepas Diubahsuai (Dimensi Pentadbiran, Penskoran dan Analisis)	159



Jadual 4.5 :	Senarai Item Sebelum dan Selepas Diubahsuai (Dimensi Interpretasi dan Tindakan Susulan)	169
Jadual 4.6 :	Pecahan Sampel Kajian Mengikut Jantina dan Major	175
Jadual 4.7 :	Skor ILP IPG Versi 2 - Dimensi Konsep Asas	185
Jadual 4.8 :	Skor ILP IPG Versi 2- Dimensi Perancangan	193
Jadual 4.9 :	Skor ILP IPG Versi 2 - Dimensi Pembinaan	200
Jadual 4.10 :	Skor ILP IPG Versi 2 – Dimensi Pentadbiran, Penskoran dan Analisis	208
Jadual 4.11:	Skor ILP IPG Versi 2 – Dimensi Interpretasi dan Tindakan Susulan	216
Jadual 4.12:	Perbandingan Dapatan Kajian ILP IPG Versi 1 dan 2	219
Jadual 4.13:	Person Fit ILP IPG Versi 2	220
Jadual 4.14 :	Analisis DIF Dimensi dalam ILP IPG	225
Jadual 5.1 :	Huraian Dimensi dalam ILP IPG	233

SENARAI RAJAH

	Halaman
Rajah 2.1: Model Pembinaan Ujian Brown (1983)	80
Rajah 2.2: Model Pembinaan Ujian Cohen dan Swerdlik (2002)	82
Rajah 2.3: Model Pembinaan Ujian McIntire Dan Miller (2007)	83
Rajah 2.4: Kerangka Konseptual Kajian	87
Rajah 3.1: Prosedur Pembinaan Instrumen ILP IPG	90
Rajah 3.2: Prosedur Persampelan Untuk Memilih Responden	95
Rajah 4.1: Kebolehpercayaan Individu dan Item ILP IPG Versi 1	118
Rajah 4.2: Peta Item ILP IPG Versi 1	120
Rajah 4.3: Polariti Item ILP IPG Versi 1- Dimensi Konsep Asas	121
Rajah 4.4: Urutan Item Mengikut Kesesuaian (Item Fit): Dimensi Konsep Asas	122
Rajah 4.5: Crossplot Item K2	123
Rajah 4.6: Analisis Distraktor bagi Item K2	124
Rajah 4.7: Kebolehpercayaan Individu dan Item ILP IPG Versi 1 - Dimensi Konsep Asas	125
Rajah 4.8: Peta Item ILP IPG Versi 1- Dimensi Konsep Asas	128
Rajah 4.9: Unidimensi ILP IPG Versi 1- Dimensi Konsep Asas	130
Rajah 4.10: Ujian penentuan piawai residual korelasi - Dimensi Konsep Asas	130
Rajah 4.11: Polariti Item ILP IPG Versi 1: Dimensi Perancangan	131
Rajah 4.12: Urutan Item Mengikut Kesesuaian (Item Fit): Dimensi Perancangan	132
Rajah 4.13: Crossplot R3	133
Rajah 4.14: Crossplot R4	133
Rajah 4.15: Kebolehpercayaan Individu dan Item ILP IPG Versi 1 - Dimensi Perancangan	135
Rajah 4.16: Peta Item ILP IPG Versi 1- Dimensi Perancangan	138
Rajah 4.17: Peta Item ILP IPG Versi 1- Dimensi Pembinaan	140
Rajah 4.18: Unidimensi ILP IPG Versi 1- Dimensi Perancangan	141

Rajah 4.19:	Ujian penentuan piawai residual korelasi - Dimensi Perancangan	142
Rajah 4.20:	Polariti Item ILP IPG Versi 1: Dimensi Pembinaan	143
Rajah 4.21:	Urutan Item Mengikut Kesesuaian (Item Fit)- Dimensi Pembinaan	144
Rajah 4.22:	Pola Respon Skalogram Guttman- Invididu Berpencapaian Rendah	144
Rajah 4.23:	Crossplot Item B10	145
Rajah 4.24:	Crossplot Item B6	146
Rajah 4.25:	Analisis Distraktor bagi Item B10 dan B6	146
Rajah 4.26:	Kebolehpercayaan Individu dan Item ILP IPG Versi 1- Dimensi Pembinaan	148
Rajah 4.27:	Peta Item ILP IPG Versi 1- Dimensi Pembinaan	151
Rajah 4.28:	Unidimensi ILP IPG Versi 1- Dimensi Pembinaan	152
Rajah 4.29:	Ujian penentuan piawai residual korelasi - Dimensi Pembinaan	153
Rajah 4.30:	Polariti Item ILP IPG Versi 1: Dimensi Pentadbiran, Penskoran dan Analisis	154
Rajah 4.31:	Urutan Item Mengikut Kesesuaian (Item Fit) : Dimensi Pentadbiran, Penskoran dan Analisis	155
Rajah 4.32:	Pola Respon Skalogram Guttman-Individu Berpencapaian Tinggi	155
Rajah 4.33:	Pola Respon Skalogram Guttman-Individu Berpencapaian Rendah	156
Rajah 4.34:	Crossplot Item A9	157
Rajah 4.35:	Crossplot Item A12	157
Rajah 4.36:	Analisis Distraktor bagi item A12 dan A9	158
Rajah 4.37:	Kebolehpercayaan Individu dan Item ILP IPG Versi 1 - Dimensi Pentadbiran, Penskoran dan Analisis	160
Rajah 4.38:	Peta Item ILP IPG Versi 1- Dimensi Pentadbiran, Penskoran dan Analisis	163



Rajah 4.39:	Unidimensi ILP IPG Versi 1-Dimensi Pentadbiran, Penskoran dan Analisis	165
Rajah 4.40:	Ujian penentuan piawai residual korelasi IPG Versi 1 - Dimensi Pentadbiran, Penskoran dan Analisis	165
Rajah 4.41:	Polariti Item ILP IPG Versi 1: Dimensi Interpretasi dan Tindakan Susulan	166
Rajah 4.42:	Infit MNSQ dan Outfit MNSQ ILP IPG Versi 1 - Dimensi Interpretasi dan Tindakan Susulan	167
Rajah 4.43:	Crossplot Item I1	168
Rajah 4.44:	Peta Item ILP IPG Versi 1- Dimensi Interpretasi dan Tindakan Susulan	170
Rajah 4.45:	Kebolehpercayaan Individu dan Item ILP IPG Versi 1 - Dimensi Interpretasi dan Tindakan Susulan	171
Rajah 4.46:	Unidimensi ILP IPG Versi 1- Dimensi Interpretasi dan Tindakan Susulan	173
Rajah 4.47:	Ujian penentuan piawai residual korelasi ILP IPG Versi 1 - Dimensi Interpretasi dan Tindakan Susulan	174
Rajah 4.48:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2	176
Rajah 4.49:	Peta Item ILP IPG Versi 2	178
Rajah 4.50:	Polariti Item ILP IPG Versi 2-Dimensi Konsep Asas	179
Rajah 4.51:	Urutan Item Mengikut Kesesuaian (Item Fit) : Dimensi Konsep Asas	180
Rajah 4.52:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2 (Konsep Asas)	181
Rajah 4.53:	Peta Item ILP IPG Versi 2- Dimensi Konsep Asas	184
Rajah 4.54:	Unidimensi ILP IPG Versi 2- Dimensi Konsep Asas	186
Rajah 4.55:	Ujian penentuan piawai residual korelasi ILP IPG Versi 2 - Dimensi Konsep Asas	186
Rajah 4.56:	Polariti Item ILP IPG Versi 2: Dimensi Perancangan	187
Rajah 4.57:	Urutan Item Mengikut Kesesuaian (Item Fit) : Dimensi Perancangan	188

Rajah 4.58:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2	189
	- Dimensi Perancangan	
Rajah 4.59:	Peta Item ILP IPG Versi 2- Dimensi Perancangan	192
Rajah 4.60:	Unidimensi ILP IPG Versi 2- Dimensi Perancangan	194
Rajah 4.61:	Ujian penentuan piawai residual korelasi ILP IPG Versi 2	194
	- Dimensi Perancangan	
Rajah 4.62:	Polariti Item ILP IPG Versi 2: Dimensi Pembinaan	195
Rajah 4.63:	Urutan Item Mengikut Kesesuaian (Item Fit): Dimensi Pembinaan	195
Rajah 4.64:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2	196
	- Dimensi Pembinaan	
Rajah 4.65:	Peta Item ILP IPG Versi 2- Dimensi Pembinaan	199
Rajah 4.66:	Unidimensi ILP IPG Versi 2- Dimensi Pembinaan	201
Rajah 4.67:	Ujian penentuan piawai residual korelasi ILP IPG Versi 2	201
	- Dimensi Pembinaan	
Rajah 4.68:	Polariti Item ILP IPG Versi 2: Dimensi Pentadbiran, Penskoran dan Analisis	202
Rajah 4.69:	Urutan Item Mengikut Kesesuaian (Item Fit) : Dimensi Pentadbiran, Penskoran dan Analisis	203
Rajah 4.70:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2- Dimensi Pentadbiran, Penskoran dan Analisis	204
Rajah 4.71:	Peta Item ILP IPG Versi 2- Dimensi Pentadbiran, Penskoran dan Analisis	207
Rajah 4.72:	Unidimensi ILP IPG Versi 2- Dimensi Pentadbiran, Penskoran dan Analisis	209
Rajah 4.73:	Ujian penentuan piawai residual korelasi ILP IPG Versi 2 - Dimensi Pentadbiran, Penskoran dan Analisis	209
Rajah 4.74:	Polariti Item ILP IPG Versi 2: Dimensi Interpretasi dan Tindakan Susulan	210
Rajah 4.75:	Urutan Item Mengikut Kesesuaian (Item Fit): Dimensi Interpretasi dan Tindakan Susulan	211

Rajah 4.76:	Kebolehpercayaan Individu dan Item ILP IPG Versi 2- Dimensi Interpretasi dan Tindakan Susulan	212
Rajah 4.77:	Peta Item ILP IPG Versi 2- Dimensi Interpretasi dan Tindakan Susulan	215
Rajah 4.78:	Unidimensi ILP IPG Versi 2- Dimensi Interpretasi dan Tindakan Susulan	217
Rajah 4.79:	Ujian penentuan piawai residual korelasi ILP IPG Versi 2 - Dimensi Interpretasi dan Tindakan Susulan	217
Rajah 4.80:	Analisis GDIF bagi Dimensi Konsep Asas	221
Rajah 4.81:	Analisis GDIF bagi Dimensi Perancangan	222
Rajah 4.82:	Analisis GDIF bagi Dimensi Pembinaan	223
Rajah 4.83:	Analisis GDIF bagi Dimensi Pentadbiran, Penskoran dan Analisis	223
Rajah 4.84:	Analisis GDIF bagi Dimensi Interpretasi dan Tindakan Susulan	224

SENARAI SINGKATAN

KPM	- Kementerian Pendidikan Malaysia
RMK 10	- Rancangan Malaysia Ke-10
TIMSS	- <i>Trends in Mathematics and Science Study</i>
PISA	- <i>Programme for International Student Assessment</i>
SPPK	- Sistem Pentaksiran Pendidikan Kebangsaan
IPG	- Institut Pendidikan Guru
PISMP	- Program Ijazah Sarjana Muda Perguruan Dengan Kepujian
STCEAS	- <i>Standards for Teacher Competence in Educational Assessment of Students</i>
PPPM	- Pelan Pembangunan Pendidikan Malaysia
PBS	- Pentaksiran Berasaskan Sekolah
TALQ	- <i>Teacher Assessment Literacy Questionnaire</i>
ALI	- <i>Assessment Literacy Inventory</i>
CALI	- <i>Classroom Assessment Literacy Inventory</i>
SGM	- Standard Guru Malaysia
KSSR	- Kurikulum Standard Sekolah Rendah
ILP IPG	- Instrumen Literasi Pentaksiran Institut Pendidikan Guru
AFT	- <i>American Federation of Teachers</i>
NCME	- <i>The National Council on Measurement in Education</i>
NEA	- <i>National Education Association</i>
CTT	- Teori Ujian Klasik
IRT	- <i>Item Response Theory</i>
DIF	- Differential Item Functioning
GDIF	- <i>Gender Differential Item Functioning</i>
JSU	- Jadual Spesifikasi Ujian

SENARAI LAMPIRAN

	Halaman
Lampiran A: Bilangan Pelajar PISMP Semester 7 Mengikut IPG Kampus	248
Lampiran B: Instrumen Literasi Pentaksiran Institut Pendidikan Guru Versi 1	249
Lampiran C: Instrumen Literasi Pentaksiran Institut Pendidikan Guru Versi 2	266
Lampiran D: Dapatan Kajian tentang Konsep Asas	284
Lampiran E: Dapatan Kajian tentang Perancangan	285
Lampiran F: Dapatan Kajian tentang Pembinaan	286
Lampiran G: Dapatan Kajian tentang Pentadbiran, Penskoran dan Analisis	287
Lampiran H: Dapatan Kajian tentang Interpretasi dan Tindakan Susulan	287
Lampiran I: Surat Kelulusan	289

RUJUKAN

- Ackerman, P. L., Bowen, K. R., Beier, M. E., dan Kanfer, R. 2001. Determinants of Individual Differences and Gender Differences in Knowledge. *Journal of Educational Psychology*, **93**(4), 797–825.
- Ackerman, P. L., Kanfer, R., dan Beier, M. E. 2013. Trait Complex, Cognitive Ability, and Domain Knowledge Predictors of Baccalaureate Success, STEM Persistence, and Gender Differences. *Journal of Educational Psychology*, **105**(3), 911–927.
- Aiken, L. R. 2006. *Psychological testing and assessment*. Boston : Allyn and Bacon.
- Airasian, P. W. 1999. *Assessment in the classroom: A concise approach* (2nd ed.). New York: McGraw-Hill Companies, Inc. <http://doi.org/10.1080/00141840701219510>
- Airasian, P. W. 2001. *Classroom assessment: Concepts and applications* (4th ed.). Boston: McGraw-Hill.
- Akbariah Mohd Mahdzir. 2009. *Penerokaan ciri-ciri psikometrik instrumen pentaksiran pemikiran kritis Malaysia dan model pemikiran kritis Malaysia*. Tesis Ph.D, Universiti Kebangsaan Malaysia, Bangi.
- Alkharusi, H. 2011a. Development and datametric properties of a scale measuring students' perceptions of the classroom assessment environment. *International Journal of Instruction*, **4**(1), 105–120.
- Alkharusi, H. 2011b. Psychometric properties of the Teacher Assessment Literacy Questionnaire for preservice teachers in Oman. In *Procedia - Social and Behavioral Sciences* (Vol. **29**, pp. 1614–1624).
- American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), dan National Education Association (NEA. 1990. Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, **9**(4), 30–32. <http://doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Anastasi, A., dan Urbina, S. 1997. *Psychological testing*. New Jersey: Prentice Hall.
- Anderson, L. W., Krathwohl, D. R. D. R., dan Bloom, B. S. 2001. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York Longman, 302. http://doi.org/10.1207/s15430421tip4104_2
- Andrich, D. 1988. *Rasch models for measurement*. California: Sage Publication.

- Arter, J. A. 2001. Learning teams for Classroom Assessment Literacy. *NASSP Bulletin*, **85**(621), 53–65.
- Arter, J. A. 2004. Assessment for learning: classroom assessment to improve student achievement and well-being. In *Measuring Up: Assessment Issues for Teachers, Counselors and Administrators* (p. 790). Austin : TX PRO-ED, Inc.
- Arter, J., dan Mc Tighe, J. 2001. *Scoring rubric in classroom*. California: Corwin Press.
- Bahagian Pendidikan Guru. 2007. Manual Kualiti Kementerian Pelajaran Malaysia. Putrajaya: Kementerian Pelajaran Malaysia.
- Bahagian Pendidikan Guru. 2009. *Standard Guru Malaysia*. Putrajaya: Kementerian Pelajaran Malaysia.
- Baird, J. J.-A. 2010. Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice*, **17**(1), 1–5. <http://doi.org/10.1080/09695940903562682>
- Bartlett, J. E., Kotrlik, J. W., dan Higgins, C. C. 2001. Organizational research: determining appropriate sample size in survey research. *Information Technology Learning*, **19**(1), 43–50.
- Black, P., dan Wiliam, D. 1998. Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, **80**(2), 139–148. <http://doi.org/10.1002/hrm>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. J., dan Krathwohl, D. R. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goal. Handbook 1: Cognitive Domain*. New York: David McKay Company, Inc.
- Bond, T. G., dan Fox, C. B. 2007. *Applying the Rasch: Fundamental measurement in the Human Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates, Inc.
- Bond, T. G., dan Fox, C. M. 2001. *Applying the Rasch Model: Fundamental measurement in the Human Sciences*. New Jersey: Lawrence Erlbaum Associates Publishers, London.
- Bond, T. G., dan Fox, C. M. B. 2015. *Applying the Rasch: Fundamental measurement in the Human Sciences*. New York: Routledge.
- Borg, W. R., Worthen, B. R., dan White, K. R. 1993. *Measurement and evaluation in the school*. New York: Longman.
- Brookhart, S. M. 1999. *The Art and Science of classroom assessment: The missing part of pedagogy*. *Higher Education Report* (Vol. **27**). Washington, DC: ERIC Higher Education and Office of Educational Research and Improvement.

- Brookhart, S. M. 2011. Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement Issues and Practice*, 30(1), 3–12.
- Brown, F. G. 1970. *Principles of educational and psychological testing*. New York, Holt, Rinehart, and Winston.
- Campbell, C., dan Collins, V. L. 2007. Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 26(1), 9–18.
- Carey, L. M. 1994. *Measuring and evaluating school learning*. Boston: Allyn dan Bacon.
- Cohen, B. H. 2008. *Explaining psychological statistics*. New Jersey : John Wiley dan Sons.
- Cohen, L., Manion, L., dan Morrison, K. 2000. *Research Methods in Education* (5th ed.). London : The Falmer Press.
- Cohen, R. J., dan Swerdlik, M. E. 2002. *Psychological testing and assessment: An introduction to test and measurement* (5th ed.). Boston: McGraw-Hill, Inc.
- Creswell, J. W. 2009. *Research design : Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, Calif. : Sage Publications.
- Creswell, J. W. 2012. *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education Inc.
- Crocker, L., dan Algina, J. 1986. *Introduction to Classical and Modern*. Michigan: Holt, Rinehart and Winston.
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Cronbach, L. J. 1990. *Essentials of Psychological Testing*. New York: Harper dan Row.
- Dodeen, H. 2004. The relationship between item parameters and item fit. *Journal of Educational Measurement*, 41, 261–270.
- Dodeen, H., dan Johanson, G. 2001. The Prevalence of Gender DIF in Survey Data. In the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001) (p. 12p.)
- Dorans, N. J., dan Holland, P. W. 1993. DIF detection and description: Mantel-Heinszel and standardization. In *Differential item functioning* (pp. 33–66. Hillsdale, NJ: Erlbaum.

Dorans, N. J., dan Kulick, E. 1986. Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. *Journal of Educational Measurement*, 23(4), 355–368. <http://doi.org/10.1111/j.1745-3984.1986.tb00255.x>

Ebel, R. L. 1979. *Essential of Education Measurement*. Englewood Cliffs, N.J.: Prentice-Hall.

Ellis, T. J., dan Levy, Y. 2010. A guide for novice researchers: Design and development research methods.

Embretson, S. E. 1999. Issues in the measurement of cognitive abilities. In S. E. Embretson dan S. L. Hershberger (Eds.), *The New Rules of Measurement: What Every Psychologist and Educator Should Know*. Mahwah, NJ: Lawrence Erlbaum and Associate.

Fan, Y. C., Wang, T. H., dan Wang, K. H. 2011. A Web-Based Model For Developing Assessment Literacy Of Secondary In-Service Teachers. *Computers & Education*, 57(2), 1727–1740.

Fisher, W. P. J. 2007. Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions*, 21(1), 1095.

Friedenberg, L. 1995. *Psychological Testing: Design , Analysis , and Use*. Boston: Allyn and Bacon.

Gall, M. D., Gall, J. P., dan Borg, W. R. 2003. *Educational Research : An Introduction* (7th ed.). Boston, Mass.: Allyn and Bacon.

Gallagher, J. D. 1998. *Classroom Assessment for Teachers*. Upper Saddle River, NJ: Merrill.

Gay, L. R., Mills, G. E., dan Airasian, P. W. 2011. *Educational Research: Competencies for Analysis and Application* (8th ed.). New Jersey: Pearson Education Inc.

Gregory, R. J. 2011. *Psychological testing: History, principles, and applications*. Boston: Allyn dan Bacon.

Gronlund, N. E. 2003. *Assessment Of Student Achievement* (7th ed.). Boston : Allyn and Bacon.

Gronlund, N. E., dan Linn, R. L. 1990. *Measurement and evaluation in teaching*. New York : Macmillan.

Hair, J. F., Black, W. C., Babin, B. J., dan Anderson, R. E. 2010. *Multivariate Data Analysis: A Global Perspective*. New Jersey: Pearson Education Inc.

- Haladyna, T. M., dan Rodriguez, M. C. 2013. *Developing and Validating Test Items*. New York: Routledge.
- Hambleton, R. K., dan Jones, R. W. 1993. Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement Issues and Practice*, 8, 253–262.
- Hambleton, R., dan Swaminathan, H. 1985. *Item Response Theory. Principles and applications*. Boston: Kluwer Nijhoff Publishing.
- Hanna, G. S., dan Dettmer, P. 2004. *Assessment for effective teaching: using context-adaptive planning*. Boston: Allyn dan Bacon.
- Holland, P. W., dan Thayer, D. T. 1988. Differential Item Performance and the Mantel-Haenszel Procedure. In *Test Validity* (pp. 129–145). <http://doi.org/10.1017/CBO9781107415324.004>
- Hulse, S. F. 2006. Test Statistics. In *Teaching Techniques in the Radiologic Sciences* (Vol. 61(2), p. 160). New York: American Society of Radiologic Technologists.
- Ibrahim Mohamed Zin. 2007. *Membina Dan Mengesahkan Instrumen Pentaksiran Kecerdasan Pelbagai*. Tesis Ph.D. Universiti Kebangsaan Malaysia.
- Institut Pendidikan Guru Malaysia. 2011. *Buku Panduan Program Ijazah Sarjana Muda Perguruan Sesi 2011/2012*. Putrajaya: Kementerian Pelajaran Malaysia.
- Institut Pendidikan Guru Malaysia. 2013. *Statistik Bilangan Pelajar Progam Ijazah Sarjana Muda Perguruan Institut Pendidikan Guru Malaysia*.
- Institut Pendidikan Guru Malaysia. 2014. *Laporan Maklumbalas Pelanggan Program Ijazah Sarjana Perguruan Ambilan Januari 2011*. Putrajaya: Kementerian Pelajaran Malaysia.
- Kaplan, R. ., dan Saccuzzo, D. . 2005. *Psychological testing: Principles, applications and issues*. Southbank, Victoria: Thomson Wadsworth.
- Kehoe, J. 1995. Basic Item Analysis for Multiple-Choice Tests. *Practical Assessment Research Evaluation*, 4(10).
- Kementerian Pelajaran Malaysia. 2011. Pengurusan dan Pengendalian Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Rendah.
- Kementerian Pelajaran Malaysia. 2012a. *Pelan Strategik Interim Kementerian Pelajaran Malaysia 2011-2020*. Putrajaya: Kementerian Pelajaran Malaysia.
- Kementerian Pelajaran Malaysia. 2012b. *Pelan Strategik Intrerim KPM 2011-2020*. Putrajaya: Kementerian Pelajaran Malaysia.

Kementerian Pelajaran Malaysia. 2013. *Pelan Pembangunan Pendidikan 2013 - 2025*. Kuala Lumpur: Kementerian Pelajaran Malaysia: Kementerian Pendidikan Malaysia. <http://doi.org/10.1016/j.tate.2010.08.007>

Kerlinger, F. N., dan Lee, H. B. 2000. *Foundation of Behavioral Research*. Fort Worth, TX: Harcourt College Publishers.

Kline, T. J. B. 2005. *Psychological Testing: A Practical Approach to Design and Evaluation*. California: Sage Publications, Inc.

Kubiszyn, T., dan Borich, G. D. 2007. *Educational Testing and Measurement: Classroom Application and Practice*. New York: John Wiley and Sons.

Lane, S., Raymond, M., Haladyna, T. ., dan Downing, S. M. 2011. *Handbook of Test Development*. New Jersey : Lawrence Erlbaum Associates.

Lawshe, C. H. 1975. A quantitative approach to content validity. *Personnel Psychology*, **28**(1), 563–575.

Lembaga Peperiksaan Malaysia. 2008. Sistem Pentaksiran Pendidikan Kebangsaan (SPPK). Kuala Lumpur: Kementerian Pelajaran Malaysia

Lembaga Peperiksaan Malaysia. 2009. Sistem Pentaksiran Pendidikan Kebangsaan. Kuala Lumpur: Kementerian Pelajaran Malaysia

Lembaga Peperiksaan Malaysia. 2010. Pelaksanaan PBS untuk penambahbaikan UPSR dan PMR mulai Tahun 1 pada 2011 dan Tingkatan 1 2012. Kuala Lumpur: Kementerian Pelajaran Malaysia

Lembaga Peperiksaan Malaysia. 2011. Transformasi Pentaksiran. Kuala Lumpur: Kementerian Pelajaran Malaysia

Linacre, J. M. 1994. Sample Size and Item Calibration Stability. *Rasch Measurement Transactions*, **7**(4), 328.

Linacre, J. M. 1999. Investigating Rating Scale Category Utility. *Journal of Outcome Measurement*, **3**(2), 103–122.

Linacre, J. M. 2002. What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transaction*, **16**(2), 878.

Linacre, J. M. 2003. Rasch Power Analysis: Size vs. Significance: Infit and Outfit Mean-Square and Standardized Chi-Square Fit Statistic. *Rasch Measurement Transactions*, **17**(1), 918.

Linacre, J. M. 2004. Test Validity and Rasch Measurement: Construct , Content , etc . *Rasch Measurement Transactions*, **18**(1), 970–971.

- Linacre, J. M. 2005. Measurement, Meaning and Morality. In *Pacific Rim Objective Measurement Symposium (PROMS) & International Symposium On Measurement & Evaluation (ISME), 2005* (pp. 1–8).
- Linacre, J. M. 2006a. Item Discrimination and Rasch - Andrich Thresholds. *Rasch Measurement Transaction*, **20**(1), 1054.
- Linacre, J. M. 2006b. Winsteps Rasch measurement. [Computer Software]. Winsteps.
- Linacre, J. M. 2009. *A User's Guide to WINSTEPS Rasch-model Computer Programs*. Chicago, Illiois: MESA Press.
- Linacre, J. M. 2013. Reliability , Separation and Strata : Percentage of Sample in Each Level. *Rasch Measurement Transactions*, **16**(3), 1399.
- Linn, R. L., dan Miller, M. D. 2005. *Measurement and Assessment in Teaching*. the University of Michigan: Prentice Hall.
- Loewenthal, K. M. 2001. *An introduction to psychological tests and scales An introduction to psychological tests and scales*. East Sussex: Psychology Press.
- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., dan Mickelson, K. 2004. Facilitating the Development of Assessment Literacy. *Educational Measurement: Issues and Practice*, **23**(2), 26–32. <http://doi.org/10.1111/j.1745-3992.2004.tb00156.x>
- Luppescu, S. 1993. DIF detection examined. *Rasch Measurement: Transactions of the Rasch Measurement SIG, AERA*, **7**(2), 285–286.
- Lynn, R., Irving, P., dan Cammock, T. 2002. Sex differences in general knowledge. *British Journal of Psychology*, **93**, 545–556. [http://doi.org/10.1016/S0160-2896\(01\)00064-2](http://doi.org/10.1016/S0160-2896(01)00064-2)
- Lynn, R., Wilberg, S., dan Margraf-Stiksrud, J. 2004. Sex differences in general knowledge in German high school students. *Personality and Individual Differences*, **37**(8), 1643–1650. <http://doi.org/10.1016/j.paid.2004.02.018>
- Manolis, C., Levin, A., dan Dahstrom, R. 1997. A generation X scale : Creation and validation. *Educational and Psychological Measurement*, **57**(4), 666–684.
- Masters, G. N., dan Keeves, J. P. 1999. *Advances in measurement in educational research and assessment*. New York: Pergamon.
- McIntire, S. A., dan Miller, L. A. 2007. *Foundation of Psychological Testing: A Practical Approach* (Second). California: Sage Publications.
- McMunn, N., McColsky, W., dan Butler, S. 2004. Building teacher capacity in classroom assessment to improve student learning. *International Journal of*, **4**(4), 25–48.

- Merbitz, C., Morris, J., dan Grip, J. C. 1989. Ordinal Scales And Foundations Of Misinference. *Archives Of Physical Medicine And Rehabilitation*, **70**, 308–312.
- Mertler, C. A. 2003. Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference? In *Mid-Western Educational Research Association, Columbus, OH (Oct. 15–18, 2003)*.
- Mertler, C. A. 2005. Secondary Teachers' Assessment Literacy: Does Classroom Experience Make A Difference? *American Secondary Education*, **33**(2), 76–93.
- Mertler, C. A. 2009. Classroom Assessment Literacy Inventory. *Understanding Assessment*, **12**(2), 13. <http://doi.org/10.1177/1365480209105575>
- Mertler, C. A., dan Campbell, C. 2005a. Application Teachers' of Assessment Concepts: Development. In *American Educational Research Association, Montreal, Quebec, Canada* (p. 27).
- Mertler, C. A., dan Campbell, C. 2005b. Measuring Teacher Knowledge and Application of Classroom Assessment Concept: Development of the Assessment Literacy Inventory. In *Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada April 11–15, 2005* (p. 27).
- Messick, S. 1990. Validity of Test Interpretation and Use. *Cognitive Therapy and Research*, **34**, 380–387. <http://doi.org/10.1007/s10608-009-9269-3>
- Mohamad Najib Abdul Ghafar. 2011. *Pembinaan dan Analisis Ujian Bilik Darjah*. Skudai: Penerbit Universiti Teknologi Malaysia.
- Mohamad Sahari Noordin. 2002. *Pengujian Dan Penaksiran Di Bilik Darjah*. Kuala Lumpur : Pusat Penyelidikan UIAM.
- Mohd Majid Konting. 2000. *Kaedah Penyelidikan Pendidikan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Murphy, K. R., dan Davidshofer, C. O. 2005. *Psychological Testing* (Sixth). New Jersey: Pearson Prentice Hall.
- Nitko, A. J. 1983. *Educational Tests and Measurement: An Introduction*. New York: Harcourt College Publishers.
- Nitko, A. J. 2001. *Educational Assessment of Students* (3rd ed). Upper Saddle River, N.J. : Merrill.
- Norlia Goolamally. 2010. *Pembinaan Dan Pemeriksaan Ciri-Ciri Psikometrik Instrumen Pemilihan Pemimpin Sekolah*. Tesis PhD, Fakulti Pendidikan, Universiti Kebangsaan Malaysia

Nunnally, J. C., dan Bernstein, I. H. 1994. *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.

Oosterhof, A. 1994. *Classroom Applications of Educational Measurement*. New York : Macmillan.

Pajares, M. F. 1992. Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62(3), 307–332. <http://doi.org/10.3102/00346543062003307>

Pallant, J. 2007. SPSS survival manual. *Journal of Advanced Nursing*, 36(3), 478–478. <http://doi.org/10.1046/j.1365-2648.2001.2027c.x>

Pallant, J., dan 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*, Mar(46(Pt 1)), 1–18.

Paterno, J. 2001. Measuring Success: A Glossary of Assessment. *Building Cathedrals: Comparison for the 21th Century*.

Plake, B. S. 1993. Teacher Assessment Literacy: Teachers' Competencies In The Educational Assessment Of Students. *Mid-Western Educational Researcher*, 6(1), 21–27.

Plake, B. S., dan Impara, J. C. 1993. Teacher Assessment Literacy : Development of Training Modules. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*.

Popham, W. J. 1990a. *Modern Educational Measurement*. Boston : Allyn dan Bacon.

Popham, W. J. 1990b. Modern Educational Measurement: A Practitioner's Perspective.

Popham, W. J. 2006. Needed: A Dose of Assessment Literacy. *Educational Leadership*, 63(6), 84–85.

Popham, W. J. 2009a. All About Assessment: A Process- Not a Test. *Educational Leadership*, April.

Popham, W. J. 2009b. Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), 4–11. <http://doi.org/10.1080/00405840802577536>

Popham, W. J. 2011. *Classroom Assessment: What Teachers Need to Know* (6th. Ed.). Boston, MA : Pearson Education, Inc.

Reckase, M. D. 1979. Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4(3), 207–230.

- Rohaya Talib. 2009. *Pembinaan dan Pengesahan Instrumen Ujian Literasi Pentaksiran*. Tesis PhD, Fakulti Pendidikan, Universiti Teknologi Malaysia.
- Schafer, W. D. 1991. Essential Assessment Skills in Professional Education of Teachers. *Educational Measurement: Issues and Practice*, 10. <http://doi.org/10.1111/j.1745-3992.1991.tb00170.x>
- Shealy, R., dan Stout, W. 1993. A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159–194. <http://doi.org/10.1007/BF02294572>
- Sherron, C. 2000. *Psychometric development of the adaptive leadership competency profile*. Doctorate Thesis. University of North Texas.
- Silverman, D. 2004. *Doing qualitative research: A practical handbook*. London: Sage Publication.
- Siti Rahayah Ariffin. 2008. *Inovasi dalam Pengukuran dan Penilaian Pendidikan*. Bangi: Fakulti Pendidikan.
- Sivakumar Alagumalai, dan Curtis, D. D. 2005. Classical Test Theory. *Applied Rasch Model Measurement: A Book Exemplar*, 1–14. http://doi.org/10.1007/1-4020-3076-2_1
- Stanley, J. C. 1971. Reliability. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 356–442). Washington , DC: American Council on Education.
- Steinmayr, R., Bergold, S., Margraf-Stiksrud, J., dan Freund, P. A. 2015. Gender differences on general knowledge tests: Are they due to Differential Item Functioning? *Intelligence*, 50, 164–174. <http://doi.org/10.1016/j.intell.2015.04.001>
- Stiggins, R. J. 1995. Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–246.
- Stiggins, R. J. 1999a. Assessment, Student Confidence, and School Success. *Phi Delta Kappan*, November, 191–198.
- Stiggins, R. J. 1999b. Learning Teams for Assessment Literacy. *Journal of Staff Development*, 20(3), 17–21.
- Stiggins, R. J. 2001. The Unfulfilled Promise of Classroom Assessment. *Educational Measurement Issues and Practice*, 20(3), 5–15.
- Stiggins, R. J. 2002. Assessment Crisis: The Absence Of Assessment For Learning. *Phi Delta Kappan International*, 1–10.

- Stiggins, R. J. 2006. Assessment for Learning: A Key to Motivation and Achievement. *Edge*, 2(2), 3–19.
- Stiggins, R. J., dan Chappuis, J. 2005. Using Student-Involved Classroom Assessment to Close. *Theory Into Practice*, 44(1), 11–18.
- Stiggins, R. J., dan Chappuis, J. 2006. What A Difference A Word Makes: Assessment "For" Learning Rather Than Assessment "Of" Learning Helps Students Succeed. *National Staff Development Council*, 27(1), 10–14.
- Stiggins, R. J., Frisbie, D. a, dan Griswold, P. a. 1989. Inside High School Grading Practices: Building a Research Agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14. <http://doi.org/10.1111/j.1745-3992.1989.tb00315.x>
- Tittle, C. K. 1994. Toward an educational psychology of assessment for teaching and learning: Theories, contexts and validation arguments. *Educational Psychologist*, 29(3), 149–162.
- Tuckman, B. W. 1999. *Conducting Educational Research* (5th ed.). Forth Worth: Harcourt Brace College.
- Van Dalen. 2002. *Memahami Penyelidikan Pendidikan. Terjemahan*. Abdul Fatah Abdul Malik, Mohd Majid Konting.Ed. Ke-4 Serdang: Penerbit Universiti Putra Malaysia.
- Vincent Pang, dan Denis Lajium. 2008. *Penilaian Dalam Pendidikan*. Kota Kinabalu: Unit Penerbitan UMS.
- Wang, T., Wang, K., dan Huang, S. 2008. Designing A Web-Based Assessment Environment For Improving Pre-Service Teacher Assessment Literacy. *Computers & Education*, 51(1), 448–462. <http://doi.org/10.1016/j.compedu.2007.06.010>
- Waugh, R. F., dan Chapman, E. S. 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, 80–99.
- Webb, N. L. 2002. Assessment Literacy in A Standards-Based Education Setting. In *Paper presented at the Annual Meeting of the American Educational Research Association, April 1-5, 2002*. New Orleans, Louisiana.
- Wiliam, D. 2010. Standardized Testing and School Accountability. *Educational Psychologist*, (452), 107–122. <http://doi.org/10.1080/00461521003703060>
- Wright, B. D. 1993. How to Set Standards. *Rasch Measurement Transactions*, 7(3), 315.

- Wright, B. D. 1999. Fundamental Measurement for Psychology. In S. E. Embretson dan S. L. Hershbender (Eds.), *The New Rules of Measurement: What Every Psychologist and Educator Should Know*.
- Wright, B. D., dan Douglas, G. A. A. 1975. Best Procedures for Sample-Free Item Analysis. *Applied Psychological Measurement*, **14**(2), 219–225.
- Wright, B. D., dan Linacre, J. M. 1989. Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, **70**(12), 857–860.
- Wright, B. D., dan Masters, G. N. 1982. *Rating Scale Analysis*. Chicago : Mesa Press.
- Wright, B. D., dan Stone, M. 1999. *Measurement Essential*. Wilmington, Delaware: Wide Range, Inc.
- Wright, B. D., dan Stone, M. 2004. *Making measures*. Chicago, IL: The Phaneron Press.
- Wright, B. D., dan Stone, M. H. 1979. *Best Test Design*. Chicago : Mesa Press.
- Zhang, Z., dan Burry-stock, J. A. 2003. Classroom Assessment Practices and Teachers ' Self-Perceived Assessment Skills. *Applied Measurement in Education*, **16**(4), 323–342.
- Zieky, M. 2002. Ensuring the fairness of licensing tests. *Clear Exam Review*, **12**(1), 20–26.