

**PEMBINAAN DAN PENGESAHAN INSTRUMEN PENTAKSIRAN  
PRESTASI STANDARD AWAL PEMBELAJARAN DAN  
PERKEMBANGAN AWAL KANAK-KANAK**

**NOR MASHITAH BINTI MOHD RADZI**

**FAKULTI PENDIDIKAN  
UNIVERSITI MALAYA  
KUALA LUMPUR**

**2017**

**PEMBINAAN DAN PENGESAHAN INSTRUMEN PENTAKSIRAN PRESTASI  
STANDARD AWAL PEMBELAJARAN DAN PERKEMBANGAN  
AWAL KANAK-KANAK**

**NOR MASHITAH BINTI MOHD RADZI**

**TESIS DISERAHKAN SEBAGAI MEMENUHI KEPERLUAN BAGI IJAZAH DOKTOR  
FALSAFAH**

**FAKULTI PENDIDIKAN  
UNIVERSITI MALAYA  
KUALA LUMPUR**

**2017**

**UNIVERSITI MALAYA**  
**PERAKUAN KEASLIAN PENULISAN**

Nama: **Nor Mashitah Binti Mohd Radzi**

No. Matrik: **PHA110026**

Nama Ijazah: **Doktor Falsafah**

Tajuk Kertas Projek/Laporan Penyelidikan/Disertasi/Tesis (“Hasil Kerja ini”):

**Pembinaan dan Pengesahan Instrumen Pentaksiran Prestasi Standard Awal Pembelajaran dan Perkembangan Awal Kanak-kanak**

Bidang Penyelidikan:

**Pendidikan Awal Kanak-kanak/ Prasekolah**

Saya dengan sesungguhnya dan sebenarnya mengaku bahawa:

- (1) Saya adalah satu-satunya pengarang/penulis Hasil Kerja ini;
- (2) Hasil Kerja ini adalah asli;
- (3) Apa-apa penggunaan mana-mana hasil kerja yang mengandungi hakcipta telah dilakukan secara urusan yang wajar dan bagi maksud yang dibenarkan dan apa-apa petikan, ekstrak, rujukan atau pengeluaran semula daripada atau kepada mana-mana hasil kerja yang mengandungi hakcipta telah dinyatakan dengan sejelasnya dan secukupnya dan satu pengiktirafan tajuk hasil kerja tersebut dan pengarang/penulisnya telah dilakukan di dalam Hasil Kerja ini;
- (4) Saya tidak mempunyai apa-apa pengetahuan sebenar atau patut semunasabahnya tahu bahawa penghasilan Hasil Kerja ini melanggar suatu hakcipta hasil kerja yang lain;
- (5) Saya dengan ini menyerahkan kesemua dan tiap-tiap hak yang terkandung di dalam hakcipta Hasil Kerja ini kepada Universiti Malaya (“UM”) yang seterusnya mula dari sekarang adalah tuan punya kepada hakcipta di dalam Hasil Kerja ini dan apa-apa pengeluaran semula atau penggunaan dalam apa jua bentuk atau dengan apa juga cara sekalipun adalah dilarang tanpa terlebih dahulu mendapat kebenaran bertulis dari UM;
- (6) Saya sedar sepenuhnya sekiranya dalam masa penghasilan Hasil Kerja ini saya telah melanggar suatu hakcipta hasil kerja yang lain sama ada dengan niat atau sebaliknya, saya boleh dikenakan tindakan undang-undang atau apa-apa tindakan lain sebagaimana yang diputuskan oleh UM.

Tandatangan Calon

Tarikh:

Diperbuat dan sesungguhnya diakui di hadapan,

Tandatangan Saksi

Tarikh:

Nama:

Jawatan:

## ABSTRAK

Kajian ini bertujuan untuk membina dan mengesahkan Instrumen Pentaksiran Prestasi, Standard Awal Pembelajaran dan Perkembangan Awal Kanak-kanak. Kerangka pembinaan instrumen Miller (2013) telah menggabungkan 10 langkah pembentukan instrumen oleh McNamara (1996), Turner (2001) dan Wortham (2012). Kajian ini mengenal pasti empat faset pengukuran iaitu item dan domain, kanak-kanak, penilai dan kebezaan kefungisian faset dengan kumpulan umur dalam enam domain pembelajaran awal kanak-kanak. Teori Generalizibiliti (Teori G) digunakan bertujuan mengenal pasti nilai kebergantungan setiap domain. Seramai 278 orang kanak-kanak berumur dua hingga empat tahun daripada lapan buah taska di Selangor telah dipilih melalui teknik pensampelan bertujuan. Analisis pengujian item-item dilakukan menggunakan perisian Facets 3.71 dan EduG 2.0. Dapatan menunjukkan enam domain pembelajaran awal kanak-kanak memenuhi semua andaian model Facets Rasch. Dapatan juga menunjukkan perbezaan yang signifikan dari segi tahap kesukaran item dan domain, tahap kebolehan kanak-kanak, tahap pentaksiran oleh penilai dan kebezaan kefungisian faset dengan kumpulan umur. Nilai kebergantungan faset-faset dapat dijelaskan mengikut Teori G. Hasil kajian ini mempunyai empat implikasi. Dari sudut pembinaan instrumen, kajian ini menyumbang kepada pembinaan instrumen yang lebih mantap melalui maklumat sumber ralat yang dijelaskan. Dari sudut metodologikal, proses pentaksiran kanak-kanak yang dilakukan lebih sistematik menggunakan model Facets-Rasch yang adil dan objektif. Dari sudut pedagogikal, kajian ini menyumbang secara langsung dengan membuka ruang kepada guru untuk menentukan tahap kandungan aktiviti yang lebih khusus dengan zon perkembangan kanak-kanak. Dari sudut reka bentuk model pemerhatian dalam kajian ini, proses pemerhatian terhadap prestasi kanak-kanak lebih jelas dan disahkan melalui dapatan empirikal, selari dengan konteks pentaksiran taska di Malaysia.

# **DEVELOPMENT AND VALIDATION OF AN INSTRUMENT FOR PERFORMANCE ASSESSMENT EARLY LEARNING STANDARDS AND EARLY CHILDHOOD DEVELOPMENT**

## **ABSTRACT**

This study aims to develop and validate the instrument of performance assessment, Early Learning Standard and Early Childhood Development. The framework development of this instrument by Miller (2013) combines 10 steps of developing instrument by McNamara (1996), Turner (2001) and Wortham (2012). The study also identified four facets of the items and domains, children, raters and the differential functionality facet with the age group of six domains. Generalizability Theory (G-Theory) used to show the dependability of each domain. A total of 278 children aged two to four years from eight kindergartens were selected through the purposive sampling techniques. Items assessment's analyzed using Facets 3.71 and EduG 2.0. The findings show six domains of early childhood development fulfill all the assumptions of the Facets-Rasch model. Four facet measurements showed significant differences in the degree of difficulty of the items and domains, the ability of the children's, the rater assessment and the differential facet functioning with the age group. The facets dependability explained by the G-Theory. This study has four implications. From the development of the instrument, this study contributes to building a more robust instrument through information sources of error that can be explained. From a methodological issues, child assessment process is done systematically for using model Facets-Rasch is obviously a fair and objective. From the pedagogical angle, this study contributes directly to open space for teachers to determine the levels of activity that are more specific and child development zone. From a design of an observation model in this study, the observation of the performance of children more clearly and have empirical findings, in line with the assessment in the context of Malaysia's kindergarten.

## **Penghargaan**

*Dengan Nama Allah Yang Maha Pemurah Lagi Maha Penyayang*

Syukur Alhamdulillah kepada Allah S.W.T kerana memberikan saya kesihatan yang cukup, masa dan kesabaran yang tinggi untuk menyiapkan kajian ini. Semangat juang dan pengorbananku kini berbaloi akhirnya. Sandaran yang kuat kepada Allah S.W.T untuk memudahkan perjalanan ini telah diakhiri dengan sempurna dengan izin dariNya. Jutaan terima kasih kepada penyelia, Prof. Madya Dr. Mariani Md. Nor di atas bantuan, bimbingan, teguran dan nasihat yang diberikan selama ini. Terima kasih juga kepada Prof. Dr. Noorlide Abu Kassim, Prof. Dr. Saedah Siraj dan Dr. Shahrir Jamaluddin di atas ilmu yang dicurahkan.

Setinggi-tinggi penghargaan dan ucapan terima kasih yang tidak terhingga kepada Kementerian Pengajian Tinggi Malaysia dan Universiti Pendidikan Sultan Idris kerana menaja pengajian ini, seterusnya pihak Jabatan Pendidikan Awal Kanak-kanak UPSI dan Pengurusan Fakulti Pembangunan Manusia UPSI kerana memudahkan pengajian ini.

Sekalung penghargaan diberikan kepada ahli keluarga, adik beradik, ipar, suamiku Azrul, anak-anakku Aqif dan Ayeesha, terima kasih yang tidak terhingga kerana memahami. Terima kasih kepada rakan-rakan PhD, Jain Chee, Mardziah, Abdullah, Ilmee, Che Mah, Nazri, Nurul Aini, Asyraf, Zanariah Ahmad, Norisin, Nor Zanariah, Ainol Hidayah, Sahrul. Tidak dilupakan rakan-rakan Master, Atiqah, Shahirah, Zaleha, Shifa yang bersama susah dan senang sepanjang perjalanan ini. Penghargaan juga kepada Geng Teknikal NoaMA yang berjasa membantu menjayakan kajian ini. Kenangan bersama sentiasa di hati selamanya. Juga tidak dilupakan Kumpulan Guru dan Kanak-kanak Taska di Shah Alam dan di Hulu Selangor, rakan-rakan PAKK UPSI AT19 yang terlibat. Juga kepada kepada semua rakan pensyarah di UPSI dan UM dan sahabat tersayang Hasyimah, Nurul Ain Hidayah dan Mohd Effendi yang selalu memberi semangat. Juga kepada semua pihak yang terlibat secara langsung dan tidak langsung dalam melaksanakan kajian ini.

Sesungguhnya, PhD ini kuhadiahkan kepada kedua ibu bakuku, Mohd Radzi Misran dan Junaidah Md. Pol.

## Jadual kandungan

Perakuan Keaslian.....	ii
Penulisan.....	ii
Abstrak.....	iii
Abstract.....	iv
Penghargaan.....	v
Senarai Gambar Rajah.....	x
Senarai Jadual.....	xi
Senarai Simbol dan Singkatan.....	xvii
Senarai Lampiran.....	xviii

### Bab 1 Pengenalan

Pendahuluan.....	1
Latar Belakang Kajian.....	1
Masalah Kajian.....	5
Kerangka Teori.....	10
Kerangka Teoritikal dan Konseptual.....	12
Tujuan Kajian.....	14
Objektif Kajian.....	14
Soalan Kajian.....	14
Kepentingan Kajian.....	15
Definisi Istilah.....	17
Standard Pentaksiran.....	17
Pembelajaran Kanak-kanak.....	18
Perkembangan Kanak-kanak.....	18
Pentaksiran Prestasi.....	19
Batasan Kajian.....	19
Rumusan.....	20

### Bab 2 Kajian Literatur

Pengenalan.....	21
Definisi Standard Awal Pembelajaran dan Perkembangan Kanak-kanak.....	21
Model Standard Awal Pembelajaran dan Perkembangan.....	23
Domain Standard Awal Pembelajaran dan Perkembangan Kanak-kanak yang Diukur.....	28
Perkembangan Kognitif.....	29
Perkembangan Bahasa.....	36
Perkembangan Fizikal.....	38
Perkembangan Kreativiti.....	39
Perkembangan Sosioemosi.....	41
Perkembangan Rohani.....	41
Teori-teori.....	43
Teori Perkembangan Piaget (1980).....	43
Teori Vygotsky (1987).....	46
Rumusan teori Piaget dan Vygotsky.....	49

Konsep Penilaian dan Pengukuran.....	49
Konsep Pentaksiran kanak-kanak.....	50
Teori Pengukuran.....	52
Teori Ujian Klasik (TUK).....	52
Teori Respons Item (TRI).....	54
Perbandingan di antara TUK dan TRI.....	55
Model Rasch.....	56
Andaian Rasch.....	56
Teori Generalizibiliti.....	66
Model Pembinaan Instrumen.....	72
Bukti-bukti Pengesanan.....	79
Bukti berdasarkan kandungan instrumen.....	80
Bukti berdasarkan proses respons.....	81
Bukti berdasarkan struktur dalaman.....	83
Pentaksiran Kepelbagaian Pembelajaran Dan Perkembangan Kanak-Kanak Menggunakan Medium Pemerhatian.....	86
Pemerhatian Pembelajaran Dan Perkembangan.....	90
Strategi Pemerhatian Perkembangan Kanak-Kanak.....	91
Kriteria Pembuktian Kepelbagaian Pembelajaran Perkembangan Kanak- Kanak Berdasarkan Prinsip Pentaksiran Taman Asuhan Kanak- Kanak.....	93
Pentaksiran Berasaskan Prestasi.....	96
Pembelajaran Autentik Dan Pentaksiran Autentik.....	98
Tujuan pentaksiran prestasi.....	99
Keperluan dan Kepentingan Pentaksiran Prestasi.....	100
Strategi Pentaksiran Berasas Prestasi.....	102
Pengelasan Dan Susun Atur Pentaksiran Berasas Prestasi.....	104
Pembentukan Konsep Dan Komponen (Pembolehubah) Pentaksiran Berasas Prestasi.....	107
Kebolehan Kanak-kanak.....	109
Pengaruh Penilai.....	109
Faset Domain.....	111
Kesukaran Item.....	111
Menghubungkan Standard dan Pentaksiran Berasaskan Prestasi.....	111
Menghubungkan Standard Kepada Pembelajaran Autentik.....	112
Menghubungkan Standard Kepada Pentaksiran Prestasi.....	112
Rumusan.....	113

### Bab 3 Metodologi Kajian

Pengenalan .....	114
Reka Bentuk Kajian.....	114
Model Pembinaan Instrumen Miller (2013).....	115
Peringkat Rekabentuk.....	116
Peringkat Pembinaan Instrumen.....	118
Peringkat Pengesahan.....	122
Responden Kajian.....	124
Prosedur Pengumpulan Data.....	125



Prosedur Penganalisisan Data.....	126
Rumusan.....	145

#### Bab 4 Peringkat Rekabentuk Dan Pembinaan Instrumen

Pengenalan.....	145
Peringkat Rekabentuk.....	145
Menentukan Kumpulan Sasaran.....	145
Menentukan Kandungan Instrumen.....	145
Mengenal Pasti Prestasi.....	146
Menentukan Skala.....	146
Mengenal pasti penilai.....	147
Kerangka pembinaan instrumen.....	148
Peringkat Pembinaan Instrumen.....	150
Mengenal pasti Standard.....	150
Mengadaptasi Instrumen.....	150
Mentaksir Instrumen.....	152
Membezakan Tahap Kebolehan Kanak-kanak.....	152
Membezakan Ketegasan Penilai.....	152
Faset Domain.....	152
Menentukan Tahap Kesukaran Item.....	152
Tujuan Membina Instrumen.....	152
Menentukan Format Instrumen.....	154
Dapatan Kajian Rintis.....	154
Nilai Kebolehpercayaan.....	155
Kesahan Konstruk.....	155
Nilai Ukuran dan kesahan kebolehan kanak-kanak.....	156
Nilai Ukuran Domain dan Kedudukan Mengikut Tahap Kesukaran.....	158
Latihan Guru.....	161
Prosedur Pengumpulan Data Latihan.....	161
Jadual Penilaian.....	161
Dapatan dan Perbincangan Kajian Data Latihan.....	163
Rumusan.....	168

#### Bab 5 Hasil Dapatan

Pengenalan.....	170
Responden Kajian.....	170
Dapatan Kajian Pengesahan.....	170
Kefungsian Kategori Skala.....	171
Laporan Pengukuran Domain.....	177
Laporan Pengukuran Kanak-kanak.....	178
Laporan Pengukuran Penilai.....	181
Laporan Pengukuran Item.....	200
Menganalisis <i>Dependability Score</i> Kanak-kanak Menggunakan Teori-G.....	238
Ketaksamaan Kefungsian Faset ( <i>Differential Faset Functioning</i> ) Penilai dan Kumpulan Umur Kanak-kanak.....	284
Rumusan.....	320

## Bab 6 Perbincangan Dan Rumusan

Pengenalan.....	321
Ringkasan Kajian .....	321
Perbincangan Dapatan Kajian.....	329
Kefungsian Kategori Skala.....	323
Laporan Pengukuran Kanak-kanak.....	325
Laporan Pengukuran Penilai.....	326
Laporan Pengukuran Item.....	329
Menganalisis Dependability score Kanak-Kanak Mengikut Domain Menggunakan Teori-G.....	331
Analisis ketaksamaan kebezaan faset (DFF).....	336
Kesimpulan Dapatan Kajian.....	337
Rumusan.....	338
Implikasi Kajian.....	339
Implikasi Metodologikal.....	340
Implikasi Pedagogikal.....	342
Cadangan Kajian.....	343
Rumusan.....	346
Rujukan.....	348
Lampiran.....	359

## Senarai Gambar Rajah

Rajah 1.1	Kerangka Teoritikal Kajian.....	11
Rajah 1.2	Kerangka Teoritikal dan Konseptual.....	13
Rajah 2.1	Proses Penilaian oleh Miller, Linn & Gronlund (2013).....	50
Rajah 2.2	Proses Merancang Instrumen oleh McNamara (1996).....	74
Rajah 2.3	Proses Pembangunan Instrumen oleh Turner (1999).....	75
Rajah 2.4	Proses Pembinaan Instrumen oleh Wortham (2012).....	76
Rajah 2.5	Komponen yang ditaksir.....	108
Rajah 3.1	Model Pembinaan Instrumen (Miller, Lovler & McIntire, 2013) .....	116
Rajah 3.2	Proses pentaksiran menggunakan penilai oleh McNamara (1996) ....	120
Rajah 4.1	Kerangka Pembinaan Instrumen.....	149
Rajah 4.2	Penambahan aspek baru pentaksiran dalam rekabentuk instrumen.....	150
Rajah 4.3	Reka bentuk instrumen pentaksiran mengikut domain.....	151
Rajah 4.4	Jumlah item bagi semua domain dalam IPHKT 1.....	151
Rajah 4.5	Jumlah item bagi semua domain dalam IPHKT 2.....	151
Rajah 4.6	Kedudukan tahap kesukaran domain dan ketegasan penilai.....	160
Rajah 5.1	Pemetaan guru penilai, kanak-kanak dan item bagi domain kognitif IPHKT 1.....	203
Rajah 5.2	Pemetaan guru penilai, kanak-kanak dan item bagi domain kognitif IPHKT 2.....	206
Rajah 5.3	Pemetaan guru penilai, kanak-kanak dan item bagi domain bahasa IPHKT 1.....	209
Rajah 5.4	Pemetaan guru penilai, kanak-kanak dan item bagi domain bahasa IPHKT 2.....	212
Rajah 5.5	Pemetaan guru penilai, kanak-kanak dan item bagi domain fizikal IPHKT 1.....	215
Rajah 5.6	Pemetaan guru penilai, kanak-kanak dan item bagi domain fizikal IPHKT 2.....	218
Rajah 5.7	Pemetaan guru penilai, kanak-kanak dan item domain kreativiti IPHKT 1.....	221
Rajah 5.8	Pemetaan guru penilai, kanak-kanak dan item domain kreativiti IPHKT 2.....	224
Rajah 5.9	Pemetaan guru penilai, kanak-kanak dan item domain sosioemosi IPHKT 1.....	227
Rajah 5.10	Pemetaan guru penilai, kanak-kanak dan item domain sosioemosi IPHKT 2.....	230
Rajah 5.11	Pemetaan guru penilai, kanak-kanak dan item bagi domain rohani IPHKT 1.....	233
Rajah 5.12	Pemetaan guru penilai, kanak-kanak dan item domain rohani IPHKT 2.....	236

## Senarai Jadual

Jadual 2.1	Perbezaan Istilah Standard Awal Mengikut Negara.....	21
Jadual 2.2	Jadual Standard Prestasi.....	28
Jadual 2.3	Perbandingan Utama di antara TUK dan TRI.....	55
Jadual 2.4	Sembilan Langkah Pembangunan Instrumen yang Efektif oleh Downing (2006) .....	77
Jadual 3.1	Tahap pengukuran, skala pengukuran dan penjelasan oleh Vygotsky.....	127
Jadual 3.2	Jadual perbandingan nilai ujian (observed) dan jangkaan ( <i>expected</i> ) oleh Linacre (2008).....	136
Jadual 4.1	Perbandingan Istilah Domain Pembelajaran kanak- kanak.....	146
Jadual 4.2	Nilai kebolehpercayaan dan indeks pengasingan bagi 4 Faset Instrumen bagi kajian rintis.....	155
Jadual 4.3	Kedudukan kebolehan kanak-kanak bagi kajian rintis.....	157
Jadual 4.4	Nilai Ukuran Perbezaan Ketegasan Penilai bagi Kajian Rintis.....	158
Jadual 4.5	Nilai Ukuran dan Kedudukan Domain Perkembangan (Susah ke mudah).....	159
Jadual 4.6	Jadual Penilaian.....	162
Jadual 4.7	Frekuensi kategori dan min ukuran bagi enam domain perkembangan.....	164
Jadual 4.8	Nilai kebolehpercayaan bagi kanak-kanak dan nilai pengasingan.....	165
Jadual 4.9	Nilai Ukuran Ketegasan, Perbezaan ketegasan dan Konsistensi Penilai.....	165
Jadual 4.10	Laporan bias kanak-kanak dan penilai.....	166
Jadual 5.1	Frekuensi kategori dan min ukuran bagi IPHKT 1.....	172
Jadual 5.2	Frekuensi kategori dan min ukuran bagi IPHKT 2.....	172
Jadual 5.3	Frekuensi kategori dan min ukuran bagi domain kognitif, bahasa dan fizikal IPHKT 1.....	172
Jadual 5.4	Frekuensi kategori dan min ukuran bagi domain kreativiti, sosioemosi dan rohani IPHKT 1.....	173
Jadual 5.5	Frekuensi kategori dan min ukuran bagi domain kognitif, bahasa dan fizikal IPHKT 2.....	173
Jadual 5.6	Frekuensi kategori dan min ukuran bagi domain kreativiti, sosioemosi dan rohani IPHKT 2.....	174
Jadual 5.7	Nilai Outfit MNSQ bagi enam konstruk pembelajaran kanak-kanak dalam IPHKT 1.....	175
Jadual 5.8	Nilai Outfit MNSQ bagi enam konstruk pembelajaran kanak-kanak dalam IPHKT 2.....	175

Jadual 5.9	Nilai ambang ( <i>threshold</i> ) bagi 5 kategori skala pengukuran bagi kesemua domain dalam IPHKT 1.....	177
Jadual 5.10	Nilai ambang ( <i>threshold</i> ) bagi 5 kategori skala pengukuran bagi kesemua domain dalam IPHKT 2.....	177
Jadual 5.11	Nilai pematuhan Unidimensionaliti Berdasarkan <i>Rasch FACETS</i> .....	177
Jadual 5.12	Nilai kebolehpercayaan bagi kanak-kanak dan nilai indeks pengasingan IPHKT 1.....	178
Jadual 5.13	Nilai kebolehpercayaan bagi kanak-kanak dan nilai indeks pengasingan IPHKT 2.....	178
Jadual 5.14	Laporan pengukuran kanak-kanak bagi enam domain pembelajaran (n=80) dengan menggunakan IPHKT 1.....	180
Jadual 5.15	Laporan pengukuran kanak-kanak bagi enam domain pembelajaran (n=77) dengan menggunakan IPHKT 2.....	180
Jadual 5.16	Nilai peratus persetujuan di antara penilai, nilai fit dan kebolehpercayaan IPHKT 1.....	182
Jadual 5.17	Nilai peratus persetujuan di antara penilai, nilai fit dan kebolehpercayaan IPHKT 2.....	182
Jadual 5.18	Laporan ukuran ketegasan penilai bagi domain kognitif IPHKT 1 mengikut urutan ukuran ketegasan.....	184
Jadual 5.19	Laporan ukuran ketegasan penilai bagi domain kognitif IPHKT 2 menurut urutan ukuran ketegasan.....	185
Jadual 5.20	Laporan ukuran ketegasan penilai bagi domain bahasa IPHKT 1 menurut urutan ukuran ketegasan.....	187
Jadual 5.21	Laporan ukuran ketegasan penilai bagi domain bahasa IPHKT 2 menurut urutan ukuran ketegasan.....	188
Jadual 5.22	Laporan ukuran ketegasan penilai bagi domain fizikal IPHKT 1 menurut urutan ukuran ketegasan.....	189
Jadual 5.23	Laporan ukuran ketegasan penilai bagi domain fizikal IPHKT 2 menurut urutan ukuran ketegasan.....	191
Jadual 5.24	Laporan ukuran ketegasan penilai bagi domain kreativiti IPHKT 1 menurut urutan ukuran ketegasan.....	192
Jadual 5.25	Laporan ukuran ketegasan penilai bagi domain kreativiti IPHKT 2 menurut urutan ukuran ketegasan.....	193
Jadual 5.26	Laporan ukuran ketegasan penilai bagi domain sosioemosi IPHKT 1 menurut urutan ukuran ketegasan.....	195
Jadual 5.27	Laporan ukuran ketegasan penilai bagi domain sosioemosi IPHKT 2 menurut urutan ukuran ketegasan.....	196
Jadual 5.28	Laporan ukuran ketegasan penilai bagi domain rohani IPHKT 1 menurut urutan ukuran ketegasan.....	197
Jadual 5.29	Laporan ukuran ketegasan penilai bagi domain rohani IPHKT 2 menurut urutan ukuran ketegasan.....	199
Jadual 5.30	Ringkasan statistik bagi faset penilai dalam mengukur domain dalam IPHKT 1.....	200
Jadual 5.31	Ringkasan statistik bagi faset penilai dalam mengukur domain dalam IPHKT 2.....	200
Jadual 5.32	Nilai kebolehpercayaan dan indeks pengasingan bagi item IPHKT 1 berdasarkan domain.....	201

Jadual 5.33	Nilai kebolehpercayaan dan indeks pengasingan bagi item IPHKT 2 berdasarkan domain Nilai kebolehpercayaan dan indeks pengasingan bagi item IPHKT 1 berdasarkan domain.....	201
Jadual 5.34	Laporan pengukuran item-item bagi domain kognitif IPHKT 1.....	204
Jadual 5.35	Laporan pengukuran item-item bagi domain kognitif IPHKT 2.....	207
Jadual 5.36	Laporan pengukuran item-item bagi domain bahasa IPHKT 1.....	210
Jadual 5.37	Laporan pengukuran item-item bagi domain bahasa IPHKT 2.....	213
Jadual 5.38	Laporan pengukuran item-item bagi domain fizikal IPHKT 1.....	216
Jadual 5.39	Laporan pengukuran item-item bagi domain fizikal IPHKT 2.....	219
Jadual 5.40	Laporan pengukuran item-item bagi domain kreativiti IPHKT 1.....	222
Jadual 5.41	Laporan pengukuran item-item bagi domain kreativiti IPHKT 2.....	225
Jadual 5.42	Laporan pengukuran item-item bagi domain sosioemosi IPHKT 1.....	228
Jadual 5.43	Laporan pengukuran item-item bagi domain sosioemosi IPHKT 2.....	231
Jadual 5.44	Laporan pengukuran item-item bagi domain rohani IPHKT 1.....	234
Jadual 5.45	Laporan pengukuran item-item bagi domain rohani IPHKT 2.....	237
Jadual 5.46	Komponen Varians Faset Penilaian untuk Domain Perkembangan Kognitif IPHKT 1.....	240
Jadual 5.47	Komponen Varians Faset Penilaian untuk Domain Perkembangan Kognitif IPHKT 2.....	241
Jadual 5.48	Komponen Varians Faset Penilaian untuk Domain Perkembangan Bahasa IPHKT 1.....	243
Jadual 5.49	Komponen Varians Penilaian untuk Domain Perkembangan Bahasa IPHKT 2.....	244
Jadual 5.50	Komponen Varians Penilaian untuk Domain Perkembangan Fizikal IPHKT 1.....	246
Jadual 5.51	Komponen Varians Faset Penilaian untuk Domain Perkembangan Fizikal IPHKT 2.....	248
Jadual 5.52	Komponen Varians Faset Penilaian untuk Domain Perkembangan Kreativiti IPHKT 1.....	249

Jadual 5.53	Komponen Varians Faset Penilaian untuk Domain Perkembangan Kreativiti IPHKT 2.....	251
Jadual 5.54	Komponen Varians Faset Penilaian Domain Perkembangan Sosioemosi IPHKT 1.....	252
Jadual 5.55	Komponen Varians Faset Penilaian untuk Domain Perkembangan Sosioemosi IPHKT 2.....	254
Jadual 5.56	Komponen Varians Penilaian untuk Domain Perkembangan Rohani IPHKT 1.....	256
Jadual 5.57	Komponen Varians Penilaian untuk Domain Perkembangan Rohani IPHKT.....	257
Jadual 5.58	<i>Dependability Score</i> Koeffisien Domain Kognitif IPHKT 1.....	258
Jadual 5.59	<i>Dependability Score</i> Koeffisien Domain Kognitif IPHKT 2.....	259
Jadual 5.60	<i>Dependability Score</i> Koeffisien Domain Bahasa IPHKT 1.....	261
Jadual 5.61	<i>Dependability Score</i> Koeffisien Domain Bahasa IPHKT 2.....	262
Jadual 5.62	<i>Dependability Score</i> Koeffisien Domain Fizikal IPHKT 1.....	263
Jadual 5.63	<i>Dependability Score</i> Koeffisien Domain Fizikal IPHKT 2.....	264
Jadual 5.64	<i>Dependability Score</i> Koeffisien Domain Kreativiti IPHKT 1.....	265
Jadual 5.65	<i>Dependability Score</i> Koeffisien Domain Kreativiti IPHKT 2.....	266
Jadual 5.66	<i>Dependability Score</i> Koeffisien Domain Sosioemosi IPHKT 1.....	267
Jadual 5.67	<i>Dependability Score</i> Koeffisien Domain Sosioemosi IPHKT 2.....	268
Jadual 5.68	<i>Dependability Score</i> Koeffisien Domain Rohani IPHKT 1.....	269
Jadual 5.69	<i>Dependability Score</i> Koeffisien Domain Rohani IPHKT 2.....	270
Jadual 5.70	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kognitif IPHKT 1.....	270
Jadual 5.71	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kognitif IPHKT 2.....	272
Jadual 5.72	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Bahasa IPHKT 1.....	273
Jadual 5.73	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Bahasa IPHKT 2.....	274

Jadual 5.74	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Fizikal IPHKT 1.....	275
Jadual 5.75	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Fizikal IPHKT 2.....	276
Jadual 5.76	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kreativiti IPHKT 1.....	277
Jadual 5.77	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kreativiti IPHKT 2.....	278
Jadual 5.78	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Sosioemosi IPHKT 1.....	279
Jadual 5.79	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Sosioemosi IPHKT 2.....	281
Jadual 5.80	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Rohani IPHKT 1.....	282
Jadual 5.81	Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Rohani IPHKT 2.....	283
Jadual 5.82	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kognitif IPHKT.....	287
Jadual 5.83	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kognitif IPHKT.....	289
Jadual 5.84	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain bahasa IPHKT 1.....	292
Jadual 5.85	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain bahasa IPHKT 2.....	295
Jadual 5.86	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain fizikal IPHKT 1.....	298
Jadual 5.87	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain fizikal IPHKT 2.....	301
Jadual 5.88	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kreativiti IPHKT 1.....	304
Jadual 5.89	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kreativiti IPHKT 2.....	307
Jadual 5.90	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain sosioemosi IPHKT 1.....	310
Jadual 5.91	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain sosioemosi IPHKT 2.....	313
Jadual 5.92	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain rohani IPHKT 1.....	316
Jadual 5.93	Laporan bias guru penilai dan kumpulan umur kanak-kanak domain rohani IPHKT 2.....	319



## **Senarai Simbol dan Singkatan**

TUK	Teori Ujian Klasikal
TRI	Teori Respons Item
FRI	Fungsi Respons Item
ICC	Keluk Ciri Item
AERA	<i>American Educational Research Education</i>
APA	<i>American Psychological Association</i>
NCME	<i>National Council on Measurement in Education</i>
PTMEA	Nilai polariti item
GT	Teori Generalizibiliti
ZPP	Zon Perkembangan Proksimal

University of Malaya

## **Senarai Lampiran**

Lampiran A	Perlantikan panel pakar kesahan
Lampiran B	Permohonan kebenaran menjalankan rakaman video penyelidikan lapangan di taska
Lampiran C	Permohonan kebenaran menjalankan kajian lapangan (rintis)
Lampiran D	Surat jemputan ke bengkel latihan guru
Lampiran E	Senarai Item Mengikut Domain IPHKT 1
Lampiran F	Senarai Item Mengikut Domain IPHKT 2
Lampiran G	Tahap Kebolehan Kanak-kanak Umur 2 dan 3 Tahun
Lampiran H	Tahap Kebolehan Kanak-kanak Umur 3 dan 4 Tahun
Lampiran I	Output Unidimensionaliti Mengikut Domain IPHKT 1
Lampiran J	Output Unidimensionaliti Mengikut Domain IPHKT 2
Lampiran K	Instrumen Latihan Guru
Lampiran L	Instrumen IPHKT 1
Lampiran M	Instrumen IPHKT 2

## **Bab 1 Pengenalan**

### **Pendahuluan**

Bab ini menjelaskan kajian yang dijalankan tentang pembinaan instrumen standard awal pembelajaran, perkembangan kanak-kanak berasaskan penilaian berasas prestasi dalam kalangan kanak-kanak. Kajian ini melibatkan kanak-kanak berumur 2 hingga 4 tahun yang berada di taman asuhan kanak-kanak (TASKA) di swasta. Bab ini membincangkan latar belakang kajian, masalah kajian, kerangka teori kajian, kerangka konsep dan teoritikal kajian, tujuan dan objektif kajian, soalan kajian, kepentingan kajian, definisi konsep dan operasional dan batasan kajian.

### **Latar Belakang Kajian**

Sistem pendidikan kini telah membawa perubahan cara berfikir para pendidik tentang penilaian kanak-kanak. Dengan berkembangnya sistem pentaksiran berasaskan kriteria dan prestasi, seseorang kanak-kanak diukur berdasarkan prestasi ditunjukkan yang selari dengan aspek perkembangan yang hendak dinilai. Ujian standard tradisional berbentuk ujian-pensil-kertas dan ujian aneka pilihan tidak dapat mengukur dengan baik sejauh mana kebolehan kanak-kanak berfikir dan menyelesaikan masalah (Wiggins, 1993) dan pentaksiran secara individu perlu dilakukan kepada peringkat awal kanak-kanak (Banerji, 1999). Pentaksiran autentik atau pentaksiran berasaskan prestasi, adalah langkah untuk mengukur kebolehan kanak-kanak berdasarkan prestasi mereka ke atas item atau tugas yang diberikan.

Terdapat kepelbagaian perbezaan perkembangan kanak-kanak yang mempengaruhi tahap prestasi mereka. Faktor demografi seperti umur, jantina dan lokasi menyumbang kepada pencapaian dalam melakukan sesuatu aktiviti. Selain faktor demografi, perbezaan prestasi boleh dilihat melalui domain-domain perkembangan. Kepelbagaian prestasi ini membawa kembali peranan pentaksiran

terhadap proses pendidikan kanak-kanak secara holistik (Morrison, 2012). Pendidikan awal kanak-kanak telah mula menjadi satu keutamaan kini kerana mereka adalah aset negara yang berharga dan wajar diberi didikan meliputi keseluruhan dimensi perkembangan iaitu kognitif, bahasa, fizikal, kreativiti, sosioemosi dan rohani.

Pada abad ke-21 ini telah berlakunya peningkatan penggunaan alat pengujian untuk mengukur pencapaian dan prestasi pada peringkat awal kanak-kanak dan pada waktu yang sama, ramai ibu bapa dan golongan profesional dalam pendidikan awal kanak-kanak mempersoalkan kurangnya penekanan terhadap alat pengujian terhadap kanak-kanak. Maka strategi pentaksiran pada peringkat awal kanak-kanak perlu memberi faedah kepada mereka dan meningkatkan proses pembelajaran (Copple & Bredekamp, 2009; Wiggins, 1993, 1998). Ia bertujuan untuk menentukan tahap perkembangan kanak-kanak dan mengenal pasti kemahiran yang belum atau perlu dikuasai. Walau bagaimanapun, tujuan pentaksiran tidak lagi mengambil kira keperluan dan minat kanak-kanak apabila mereka melangkah ke alam persekolahan (Wortham, 2012). Prinsip pentaksiran untuk peringkat kanak-kanak bukan sahaja berkaitan dengan penilaian semata-mata, tetapi memberi implikasi kepada program penilaian dan kualiti (Epstein, 2004). Pada peringkat awal kanak-kanak, pentaksiran terhadap perkembangan mereka adalah fokus utama.

Pada peringkat umur memasuki taska, pentaksiran dilakukan untuk menjangkakan tahap perkembangan pembelajarannya. Pentaksiran dapat membantu guru mengetahui perkembangan kanak-kanak. Kanak-kanak tidak dapat melihat dan memahami kemajuan perkembangan diri mereka. Rekod maklumat kanak-kanak adalah langkah asas untuk guru memahami perkembangan sebenar kanak-kanak (Darragh, 2009).

Maka, berdasarkan hasil pentaksiran, dapat menentukan apakah langkah yang terbaik menyelesaikan masalah tersebut (Greenspan,1996; Wodrich, 1984) dalam Wortham (2012). Apabila masalah dikenal pasti, perancangan perlu dilakukan untuk membantu kanak-kanak menyelesaikan masalah itu dalam tempoh sebelum memasuki prasekolah.

Pada tahun 2010, modul pentaksiran perkembangan murid di prasekolah telah dihasilkan oleh Bahagian Pembangunan Kurikulum, Kementerian Pelajaran Malaysia. Pendidikan prasekolah di Malaysia bertujuan memperkembangkan potensi kanak-kanak berumur 4 hingga enam tahun secara menyeluruh dan bersepadu dalam aspek jasmani, emosi, rohani, intelek dan sosial melalui persekitaran pembelajaran yang selamat, menyuburkan serta aktiviti yang menyeronokkan, kreatif dan bermakna. Ini adalah untuk meningkatkan kemahiran dan keyakinan dalam diri kanak-kanak agar berjaya dalam persekitaran sedia ada dan bersedia memikul cabaran dan tanggungjawab di sekolah rendah kelak. Pentaksiran yang digunakan berasaskan formatif dan autentik melalui pengajaran dan pembelajaran harian.

Pelaksanaan proses pembelajaran bertujuan mencapai objektif pengajaran dan hasil pembelajaran yang dikuasai merupakan tanda aras penguasaan kanak-kanak. Sejauhmana penguasaan yang diperolehi? Pengesananannya memerlukan pentaksiran khusus kerana kumpulan sasaran adalah kanak-kanak. Penilaian yang berkesan menggunakan proses pemerhatian yang sistematik yang membawa tujuan khusus iaitu menentukan perkembangan kognitif, bahasa, fizikal, kreativiti, sosioemosi dan rohani, mengenal pasti minat dan gaya belajar kanak-kanak, membuat perancangan, mengesan keperluan individu kanak-kanak, menentukan kadar kemajuan kanak-kanak, menyediakan maklumat kepada ibu bapa dan golongan profesional untuk bimbingan selanjutnya (Morrison, 2012).

Ahli psikometrik menekankan isu kebolehpercayaan dan kesahan semasa menggunakan pentaksiran berasaskan prestasi (Green, 1995). Pentaksiran autentik atau pentaksiran berasaskan prestasi menggunakan strategi pengukuran dengan membenarkan kanak-kanak menunjukkan kefahaman dan respon terhadap sesuatu konsep atau penguasaan sesuatu kemahiran. Soal jawab dengan guru atau pengasuh, tugas, aktiviti-aktiviti yang dilakukan yang mendedahkan tahap kebolehan kanak-kanak akan didokumentasikan sebagai hasil kerja atau item prestasi (Wortham, 2012) dan dinilai dengan menggunakan skala rating oleh penilai. Bagaimana kebolehan seseorang kanak-kanak itu dapat diukur secara objektif dan tepat jika membuat penilaian secara subjektif, dan berpanjangan yang melibatkan lebih daripada seorang penilai? Penilai mempunyai pengalaman yang tersendiri, kepakaran dan berkemungkinan menjadi bias. Satu pendekatan yang sistematik diperlukan semasa proses penilaian kanak-kanak menggunakan penilai untuk menyediakan kualiti psikometrik yang merupakan asas kepada kebolehpercayaan dan kesahan yang tinggi.

Model pengukuran Rasch, diperkenalkan khusus oleh Rasch (1960), ada mengambil kira kegunaan pendekatan pentaksiran berasaskan prestasi kerana ciri-ciri pengukuran secara objektif. Model Pengukuran Rasch Pelbagai Facet (PRPF) dibangunkan oleh Linacre (1994) lanjutan daripada model Rasch iaitu, daripada dua komponen iaitu *person ability* dan *item difficulty* kepada empat komponen atau lebih iaitu *person ability*, *rater effect*, domain dan *item difficulty*. Fokus kajian ini adalah membina dan mengesahkan instrumen pentaksiran prestasi standard awal pembelajaran dan perkembangan awal kanak-kanak taska berasaskan penilaian prestasi untuk mengenal pasti empat komponen pengukuran iaitu faset kebolehan kanak-kanak, faset penilai, faset domain dan faset tahap kesukaran item menggunakan Model Pengukuran Rasch Pelbagai Facet (PRPF) dan Teori Generalizibiliti.

## **Masalah Kajian**

Pentaksiran kanak-kanak taska mesti meliputi keseluruhan aspek perkembangan kanak-kanak seperti kognitif, bahasa, kreativiti, fizikal, sosioemosi dan rohani. Perkara paling utama, pentaksiran sebagai alat refleksi dan maklumat ini boleh mengenal pasti dengan mendalam aspek potensi kanak-kanak yang terlindung, bahagian pengajaran yang sukar atau mudah bagi kanak-kanak dan mengetahui apa yang perlu dicapai pada usianya. Amalan refleksi adalah aspek penting dalam bidang kanak-kanak dan ia juga mampu meningkatkan tahap pembangunan profesionalisme pendidik dalam bidang ini. Walau bagaimanapun, amalan refleksi oleh pengamal dalam bidang kanak-kanak pada abad ke-21 ini semakin tenggelam sedangkan ia sangat penting dan sebenarnya merupakan kunci penghubung bidang kanak-kanak dengan pelbagai pihak ketiga yang berkaitan dengan kanak-kanak (Alice & Anna, 2008; Chorney, 2006; Paige-Smith & Craft, 2009).

Kementerian Pelajaran Malaysia telah menetapkan satu standard pentaksiran murid prasekolah selari dengan Kurikulum Standard Prasekolah Kebangsaan yang dilaksanakan di semua prasekolah mulai 2010. Standard pentaksiran ini merupakan satu sistem untuk guru mengumpul maklumat tentang perkembangan murid dan bertujuan menjalankan penilaian terhadap mereka. Hasil pentaksiran ini membolehkan guru memperkembangkan potensi murid sepenuhnya dalam domain kognitif, afektif dan psikomotor. Walau bagaimanapun, proses pentaksiran ini tidak dapat dimulakan pada umur sebelum memasuki prasekolah iaitu semasa di peringkat TASKA kerana ketiadaan instrumen pentaksiran dan panduan pentaksiran yang sesuai dengan perkembangan awal bagi kanak-kanak yang berumur 4 tahun dan ke bawah. Oleh itu, pengesanan awal berdasarkan rekod pentaksiran semasa TASKA perlu diwujudkan dengan cara yang sistematik.

Sekiranya pentaksiran tidak dilakukan pada peringkat ini, tahap kesediaan kanak-kanak tidak dapat diketahui dan seterusnya menggagalkan usaha Falsafah Pendidikan Negara dalam Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan untuk melahirkan insan yang berilmu dan berakhlak, seimbang dan harmonis, yang boleh mencapai kesejahteraan diri dan memberi sumbangan kepada keharmonian dan kemakmuran masyarakat dan negara. Ketiadaan instrumen pentaksiran ini juga menjejaskan matlamat penting sistem pendidikan kebangsaan dan misi memastikan perkembangan emosi dan kognitif kanak-kanak lebih optimum dan bersedia apabila memasuki alam persekolahan formal di umur lima tahun.

Mengikut statistik 2007, populasi kanak-kanak yang berumur di bawah 4 tahun adalah seramai 3,116,200 orang. Dari jumlah ini hanya 10% atau seramai 300,000 orang kanak-kanak yang dianggarkan menghadiri taman asuhan kanak-kanak di 8,841 buah TASKA (iaitu 2,176 buah TASKA yang berdaftar dengan Jabatan Kebajikan Masyarakat dan di 6,665 buah TASKA yang tidak berdaftar dengan Jabatan Kebajikan Masyarakat). Dianggarkan juga daripada 8,841 buah TASKA ini hanya 30% atau 2,653 buah TASKA (berdaftar dan tidak berdaftar) terletak di kawasan luar bandar yang memberikan perkhidmatan pengasuhan kepada 80,000 orang kanak-kanak bawah umur 4 tahun. Daripada jumlah 2,653 buah TASKA ini, sebanyak 304 TASKA adalah di bawah kelolaan KEMAS dengan bilangan kanak-kanak seramai 3,654 orang.

Berdasarkan kepada statistik di atas, dianggarkan hanya 10% jumlah kanak-kanak yang boleh menjalani proses pentaksiran untuk dilihat tahap perkembangan mereka secara sistematik. Sehubungan itu, usaha penting untuk mewujudkan instrumen pentaksiran dan panduan pentaksiran kanak-kanak di TASKA yang mesra pengguna dalam kalangan pendidik dan pengasuh perlu dimulakan dengan segera dan menepati ciri-ciri psikometrik dengan mempunyai aspek kebolehpercayaan dan kesahan yang tinggi. Proses pentaksiran kanak-kanak memerlukan satu instrumen yang berupaya



meliputi kesemua aspek perkembangan yang terdapat dalam skop kurikulum kebangsaan dan ditadbir berasaskan pemerhatian dan berdasarkan prestasi yang ditunjukkan.

Pentaksiran berasas prestasi banyak digunakan dalam penilaian peringkat awal kanak-kanak kerana kanak-kanak menunjukkan sesuatu kemahiran dalam jangka masa yang terhad dan pentaksiran boleh dilakukan secara pemerhatian (Russell, 2012). Terdapat banyak aspek seperti tahap kebolehan kanak-kanak, kesukaran item dan domain, ketegasan penilai (*rater severity*), dan skor menggunakan skala rating yang perlu dipertimbangkan oleh guru atau penilai semasa menilai kebolehan kanak-kanak ini. Perkaitan semua aspek yang relevan ini mempengaruhi ketepatan terhadap pengukuran prestasi kanak-kanak yang dinilai oleh guru mereka sendiri. Walau bagaimanapun, penilai yang terdiri daripada guru-guru TASKA masih kekurangan latihan dan panduan pentaksiran perkembangan kanak-kanak berdasarkan domain pembelajaran terutamanya melibatkan jumlah kanak-kanak yang ramai. Kekurangan latihan juga meliputi ketiadaan alat pengukuran yang sesuai digunakan untuk menganalisis data perkembangan kanak-kanak ini agar maklumat lebih tersusun secara objektif.

Prestasi kanak-kanak dinilai melalui respon yang diberikan ke atas sesuatu item. Oleh itu, kemungkinan akan berlaku berat sebelah semasa pentaksiran kerana strategi pemberian skor adalah bergantung kepada penilai. Berat sebelah atau *bias* diistilahkan sebagai kesan penilaian yang tidak adil, yang merujuk kepada kepelbagaian perbezaan yang timbul dalam penilaian prestasi yang sama oleh penilai yang berbeza (Kim, Park & Kang, 2012; Myford, & Wolfe, 2003; Saal, Downey, & Lahey, 1980; Wolfe, 2004). Ia disokong oleh beberapa kajian yang telah menyenaraikan aspek ketegasan penilai dan konsistensi pentaksiran adalah dua aspek utama kesan pentaksiran yang tidak adil sehingga mengganggu kesahan dan

kebolehpercayaan skor (Lane & Stone, 2006; Myford, & Wolfe, 2003; Saal, Downey, & Lahey, 1980; Wolfe, 2004).

Selari dengan pelbagai aspek dan masalah konsistensi pentaksiran oleh penilai, kajian ini akan menekankan penganalisan data menggunakan Model Pengukuran Rasch Pelbagai Faset (PRPF; Linacre, 2007) dan Teori Generalizabiliti (G-Theory; Brennan, 1983) yang berkebolehan mengesan kesan penilai (*rater effect*) bersama-sama aspek pengukuran yang lain dikenali sebagai faset dan kedua-duanya melengkap satu sama lain (Lynch & McNamara, 1998; Sudweeks, Reeve, & Bradshaw, 2004). Oleh itu, kesahan dan kebolehpercayaan instrumen menggunakan penilaian berasaskan prestasi dapat ditingkatkan dan merumuskan kebolehan sampel yang akan dicatatkan dengan tepat menggunakan PRPF (Engelhard, 1998) dan Teori Generalizabiliti.

*National Association for the Education of Young Children (NAEYC) 1991*, mendefinisikan pentaksiran perkembangan kanak-kanak sebagai proses mendokumentasi apa yang dilakukan oleh kanak-kanak dan bagaimana mereka melakukannya bertujuan membuat keputusan-keputusan berkenaan kanak-kanak. Pentaksiran terhadap kanak-kanak yang berasaskan prestasi perlu menentukan terlebih dahulu tujuan khusus samada berbentuk formatif atau sumatif kerana skor yang diberi penilai akan bergantung kepada bentuk penilaian tersebut. Penilaian berbentuk formatif adalah memfokuskan maklum balas kepada kekuatan dan kelemahan, oleh itu kriteria skor pemarkahan perlu mengkhususkan skor tertentu kepada aspek tertentu dan menyediakan maklumat mengenai kebolehan dan kelemahan seseorang kanak-kanak itu (Russell, 2012). Maka setiap hasil kerja berdasarkan prestasi yang ditunjukkan akan didokumentasi untuk dinilai oleh guru.

Walaupun bagaimanapun, pendidik yang berperanan sebagai penilai menghadapi kekangan menjalankan penilaian kepada hasil kerja kanak-kanak kerana latihan yang terbatas cara guna kaedah penilaian portfolio, bilangan kanak-kanak terlalu ramai, masa yang terhad dan limitasi portfolio dalam menggambarkan perkembangan kanak-kanak secara keseluruhan (Surayah, 2003). Sehubungan dengan itu, kajian ini berupaya merekod perkembangan kanak-kanak dalam satu jangka masa, memberi maklumat dan panduan yang berkesan kepada guru taska mengenai pengajaran dan penilaian hasil kerja, menentukan jumlah kanak-kanak dalam satu-satu masa penilaian dan meliputi kesemua enam bidang pembelajaran yang terkandung dalam skop Kurikulum Kebangsaan Dasar Asuhan Kanak-kanak dan Didikan secara sistematik.

Seterusnya, terdapat faktor yang mempengaruhi kesahan dan kebolehpercayaan dalam reka bentuk ujian standard. Walaupun langkah-langkah dan prosedur yang digunakan untuk memastikan kesahan dan kebolehpercayaan dalam ujian standard, terdapat faktor-faktor lain yang boleh mempengaruhi hasil pentaksiran. Beberapa faktor adalah seperti kebolehan membaca, keadaan persekitaran fizikal, ingatan atau memori, dan keadaan fizikal individu yang ditaksir (Wortham, 2012). Oleh itu, jika persekitaran pentaksiran dijalankan adalah tidak selesa, panas atau pelajar tidak cukup tidur malam sebelum ujian, skor akan terjejas. Kekurangan pematuhan kepada had masa dan ketidaktekalan dalam arahan ujian juga mempengaruhi skor ujian. Faktor-faktor lain adalah seperti tidak konsisten dalam pemberian skor rating kepada individu dan meneka jawapan ujian (Payne, 1997).

Kesahan pula dipengaruhi oleh faktor-faktor seperti arahan tidak jelas, kesukaran membaca perbendaharaan kata semasa pentaksiran, dan item ujian yang tidak sesuai untuk objektif ujian (Linn & Gronlund, 2000). Kebolehpercayaan dipengaruhi oleh bilangan item prestasi atau tempoh pentaksiran, nilai kebolehpercayaan antara penilai yang rendah, dan peristiwa lain yang memberi kesan

kepada keadaan (Linn & Gronlund, 2000; McMillan, 2007). Terdapat faktor-faktor lain mempengaruhi kesilapan yang mungkin pada alat pentaksiran dan kualiti pentaksiran itu. Oleh yang demikian, perubahan dalam kualiti pentaksiran yang disebabkan pelbagai faktor ini akan diambil kira dalam ralat piawai pengukuran (*standard error of measurement*) yang akan dianalisis melalui model pengukuran yang telah ditetapkan oleh Teori Generalizibiliti dengan menggunakan perisian EduG.

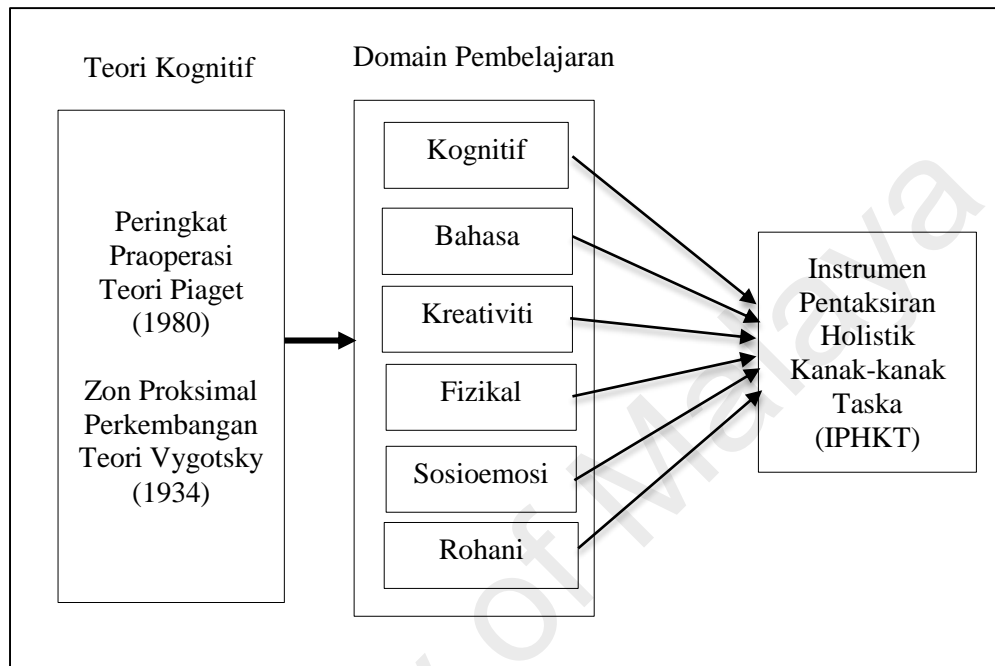
### **Kerangka Teori**

Kajian ini berlandaskan Teori Kognitif Piaget (Peringkat Praoperasi) dan Teori Vygotsky (Zon Proksimal Perkembangan). Kajian ini akan menghubungkan pembelajaran dan perkembangan kanak melalui penekanan kepada konsep kanak-kanak adalah individu yang dinamik dan berbeza tahap kebolehan.

Domain pembelajaran kanak-kanak melibatkan interaksi dinamik dengan semua aspek perkembangannya iaitu kognitif, bahasa, fizikal, kreativiti, sosioemosi, dan rohani. Melihat kepada perkembangan kognitif kanak-kanak, ia turut dipengaruhi oleh domain perkembangan yang lain. Kanak-kanak berada dalam satu peringkat yang dipengaruhi dan berubah melalui persekitaran, pengalaman mereka, pemakanan dan kesihatan, dan latar belakang genetik, dan semua ini juga mempengaruhi cara mereka berfikir. Bermula dengan berfikir, kanak-kanak merancang untuk mencapai sesuatu sasaran. Bagaimanapun, untuk mencapai sesuatu memerlukan kebolehan untuk mengambil tindakan. Tanpa kebolehan ini, apa yang difikirkan dan dirancang tidak mempunyai makna. Maka, setiap tindakan yang terhasil mempunyai satu proses yang saling berkaitan dengan semua aspek perkembangan memainkan peranan yang boleh menghubungkan antara pemikiran dan tindakan kanak-kanak.

Rajah 1.1 menunjukkan gabungan dua Teori Kognitif seperti Peringkat Praoperasi (Piaget, 1980) dan Zon Proksimal Perkembangan (Vygotsky, 1934) yang membentuk enam domain standard awal pembelajaran kanak-kanak seterusnya

menggunakan model pembinaan instrumen oleh Miller et al. (2013). Ciri-ciri psikometrik domain-domain dalam instrumen yang dibina akan diuji melalui beberapa analisis dalam model Pengukuran Rasch Pelbagai Faset (PRPF) dan Teori Generalizibiliti (EduG).



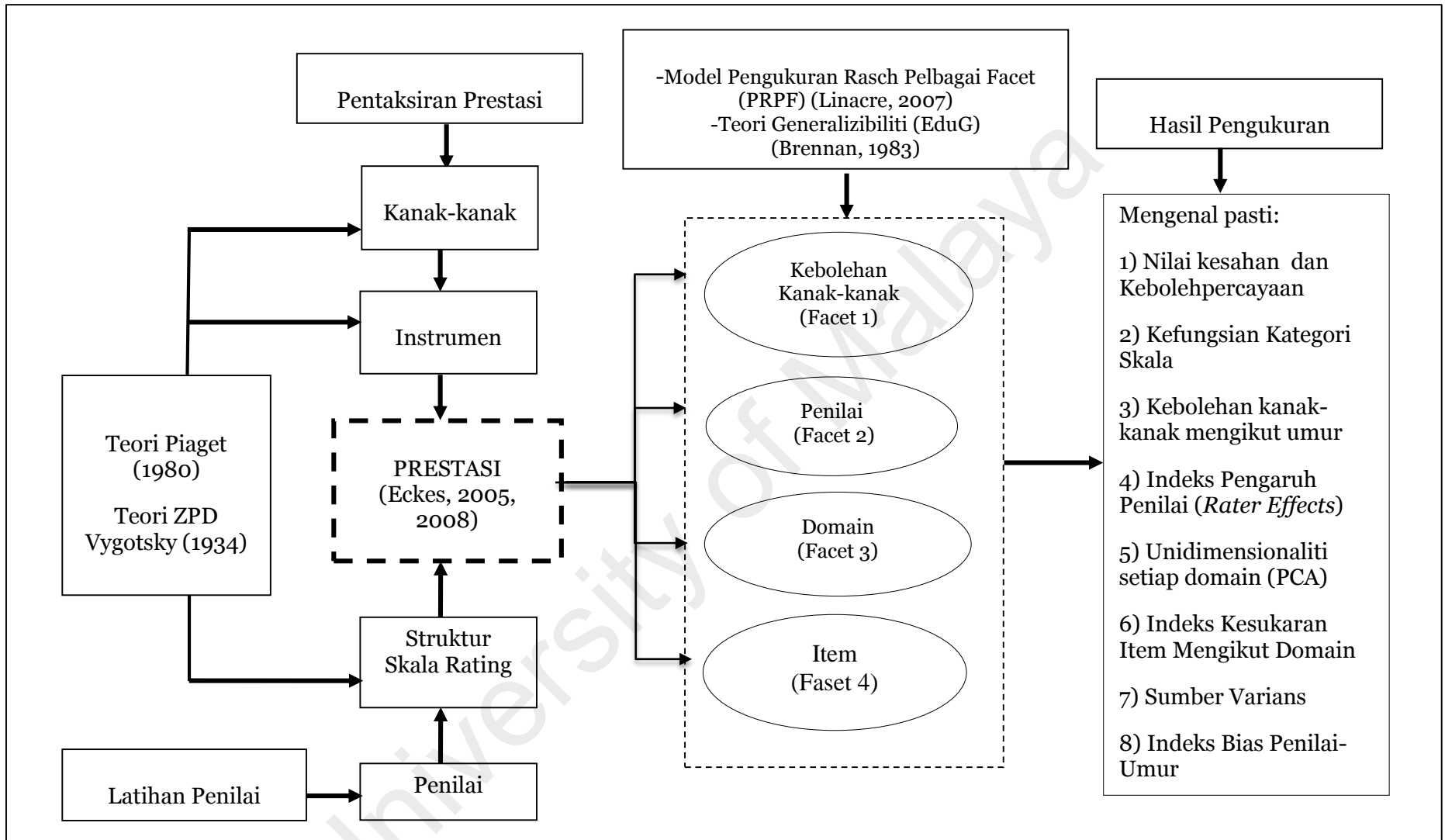
Rajah 1.1. Kerangka Teoritikal Kajian

Setiap satu teori yang berikut mempunyai konsep tersendiri dan penting dalam membawa hala tuju perbincangan sepenuhnya mengenai perkaitan teori pembelajaran dan perkembangan kanak-kanak dalam kajian ini. Teori Piaget mengenai perkembangan kognitif iaitu menjelaskan bagaimana kanak-kanak berfikir, memahami dan belajar. Vygotsky pula percaya perkembangan mental, bahasa dan sosial kanak-kanak disokong dan dikembangkan melalui interaksi sosial. Beliau menekankan pembelajaran adalah suatu aktiviti yang membangkitkan proses kepelbagaian perkembangan dan boleh berfungsi hanya apabila wujudnya interaksi kanak-kanak dengan individu dalam persekitarannya dan berkolaborasi dengan rakan-rakannya. Pencapaian kanak-kanak dalam pelbagai perkembangan dipengaruhi dengan interaksi sosial dan pada ketika interaksi itu, berlakunya proses perkembangan.

Oleh yang demikian, untuk melihat perkembangan kanak-kanak mengikut teori ini, pentaksiran perlu dilakukan semasa proses pembelajaran dimana kanak-kanak berinteraksi dengan guru dan rakan-rakan melakukan sesuatu aktiviti. Pentaksiran domain pembelajaran dapat dicatat dengan lebih maksimum. Penjelasan lanjut diberikan pada bahagian definisi konsep dan operasi.

### **Kerangka Teoritikal dan Konseptual**

Rajah 1.2 menunjukkan kerangka teoritikal dan konseptual kajian. Kanak-kanak berumur dua hingga empat tahun terlibat dalam kajian ini. Kedua-dua instrumen mengandungi enam domain pembelajaran dan perkembangan iaitu kognitif(K), bahasa(B), kreativiti(KR), fizikal(F), sosioemosi(S) dan rohani(R). Penilai akan diberi latihan dan mereka akan memerhati aktiviti kanak-kanak seterusnya memberi skor menggunakan skala rating. Keseluruhan item adalah berbentuk prestasi di mana responden iaitu kanak-kanak menghasilkan respon dan bukan memilih respon. Teori pengujian menggunakan model pengukuran Rasch Pelbagai Faset (Linacre, 2007) dan Teori Generalizibiliti EduG (Brennan, 1983) digunapakai untuk mengesan empat komponen iaitu faset kanak-kanak, faset penilai, faset domain dan faset item, seterusnya menganalisa data untuk mengenal pasti kesesuaian model pengukuran sebagai panduan pentaksiran yang berkesan. Teori Generalizabiliti melalui perisian EduG juga digunakan untuk menjelaskan lagi sumber varians bagi setiap faset dan menjadi pelengkap kepada keseluruhan kerangka teoritikal dan konseptual ini.



Rajah 1.2. Kerangka Teoritikal dan Konseptual

## **Tujuan Kajian**

Kajian ini bertujuan membina Instrumen Pentaksiran Holistik Kanak-kanak Taska (IPHKT) berasaskan pentaksiran prestasi standard awal pembelajaran dan perkembangan awal kanak-kanak. Proses pengesahan dibuat dengan menguji ciri-ciri psikometriknya menggunakan analisis model pengukuran Rasch pelbagai faset (PRPF) dan Teori Generalizibiliti (Edu G). Selain itu, rubrik instrumen pentaksiran ini dibina berdasarkan Teori Vygotsky untuk melihat tahap perkembangan enam domain iaitu kognitif, bahasa, kreativiti, fizikal, sosioemosi dan rohani. Kajian ini turut mengkaji empat pengukuran faset iaitu kanak-kanak, penilai, domain dan item bagi kumpulan umur kanak-kanak berumur dua hingga empat tahun.

## **Objektif Kajian**

- 1) Membina IPHKT berasaskan pentaksiran prestasi standard awal pembelajaran dan perkembangan awal kanak-kanak menggunakan model Miller (2013).
- 2) Menguji ciri-ciri psikometrik instrumen IPHKT yang dibina menggunakan model pengukuran Rasch pelbagai facet (PRPF) dan Teori Generalizibiliti (Edu-G) meliputi empat pengukuran faset iaitu faset kanak-kanak; faset penilai; faset domain dan faset item.

## **Soalan Kajian**

- 1) Adakah model Miller boleh digunakan untuk membina instrumen IPHKT berasaskan pentaksiran prestasi standard awal pembelajaran dan perkembangan awal kanak-kanak?



2) Adakah instrumen IPHKT yang dibina mempunyai ciri-ciri psikometrik yang baik menggunakan model Rasch pelbagai Faset dan Teori Generalizibiliti (Edu-G) meliputi empat pengukuran faset iaitu faset kanak-kanak; faset penilai; faset domain dan faset item.

### **Kepentingan Kajian**

Kajian ini berfokuskan kepada pembinaan dan pengesahan instrumen pentaksiran prestasi, standard awal pembelajaran dan perkembangan awal kanak-kanak dengan menggunakan teori perkembangan Piaget dan Vygotsky untuk menilai perkembangan kanak-kanak taska berumur dua hingga empat tahun. Pendidik di taska juga berperanan sebagai penilai terhadap prestasi kanak-kanak sebagai refleksi pengajaran. Bagaimana instrumen dapat digunakan untuk menilai perkembangan kanak-kanak oleh beberapa orang pendidik dalam masa yang sama yang mempunyai pengalaman dan deskriptif yang berbeza? Bagaimanakah hasil penilaian pendidik kepada bidang perkembangannya dapat memberikan keputusan yang adil kepada kanak-kanak tersebut? Kekurangan instrumen yang memenuhi semua keperluan tersebut mendorong pengkaji melakukan pengujian kepada instrumen penilaian yang diadaptasi dan ditambah selari dengan objektif kurikulum dan mencipta suasana persekitaran penilaian yang lebih sistematik serta memperkenalkan pendekatan model pengukuran yang berupaya menganalisa data daripada proses pemerhatian kepada skor yang lebih adil dan tepat seterusnya menghasilkan panduan penilaian yang lebih berkesan dalam pendidikan awal kanak-kanak.

Sebelum instrumen digunakan, kesahan dan kebolehpercayaan ditentukan dan dilaporkan terlebih dahulu. Instrumen ini akan menguji prestasi kanak-kanak dua hingga empat tahun dalam enam bidang perkembangan yang termaktub dalam kurikulum. Prestasi mereka ditunjukkan dan direkod sebagai hasil kerja dalam masa sama proses pengajaran dan pemerhatian oleh penilai dilakukan. Pentaksiran prestasi

pada peringkat kanak-kanak taska dinilai dalam persekitaran semulajadi mereka yang suka bermain. Walaupun bermain sambil belajar adalah penting, tetapi tidak dinafikan bahawa mempunyai instrumen yang benar-benar mampu menilai potensi dan perkembangan sebenar kanak-kanak berasaskan prestasi lebih mendesak supaya hasil pentaksiran yang diperolehi lebih tepat, sah dan boleh dipercayai serta memiliki kualiti psikometrik yang tinggi. Berdasarkan kerangka konseptual yang dicadangkan, standard psikometrik akan dapat dicapai.

Konsep pengukuran pentaksiran dibincangkan untuk mengetahui tahap pencapaian kanak-kanak. Kebiasaannya pentaksiran dilakukan secara pemerhatian terhadap kanak-kanak dan dilakukan dengan pelbagai cara oleh guru kerana sentiasa berhubungan dengan kanak-kanak semasa menjalankan aktiviti. Walaupun taska-taska di Malaysia mempunyai kemudahan yang cukup dan selesa kepada pembelajaran kanak-kanak, namun suatu kaedah pengukuran yang mantap dan menyeluruh khusus kepada sejauh mana pencapaian perkembangan kanak-kanak di peringkat taska belum lagi diwujudkan. Apabila instrumen pentaksiran perkembangan kanak-kanak ini diuji, diharapkan model pengukuran yang digunakan dapat memenuhi dan melengkapkan proses pengajaran dan pembelajaran yang diakhiri dengan proses penilaian secara pemerhatian yang sistematik, teknik penskoran yang mengukur secara objektif, berdasarkan hasil kerja yang dikumpulkan atau berbentuk portfolio dan menganalisa dengan tepat kebolehan kanak-kanak, kesukaran item dan pengaruh penilai (*rater*). Sangat kurang kajian empirikal tentang tahap perkembangan kanak-kanak di taska khususnya di taska swasta. Justeru, hasil kajian ini akan dapat memberi input berguna berkenaan proses penilaian perkembangan kanak-kanak berasaskan prestasi dan memberi impak positif kepada pihak taska dan pihak-pihak yang terlibat dalam pelaksanaan dasar pengukuran dan penilaian di institusi tertentu.

## **Definisi Istilah**

Definisi istilah yang digunakan dalam kajian ini adalah seperti berikut:

**Standard Pentaksiran.** Menurut Chapman & King (2012) standard adalah tanda aras atau kemahiran khusus yang dikenal pasti untuk dikuasai dalam bidang tertentu untuk tahap dan gred yang khusus. Tanda aras ini digunakan di peringkat persekutuan, negeri atau daerah dan kebiasaan di peringkat negeri akan menggunakannya agar selari dengan standard kebangsaan. Manakala dalam kajian ini, ia merujuk kepada standard instrumen *Early Learning and Development Standards (ELDS)* yang dibangunkan oleh Bahagian Pembangunan Kurikulum (BPK), Kementerian Pendidikan Malaysia dengan kerjasama UNICEF yang memfokuskan enam bidang perkembangan kanak-kanak berumur 2 hingga 4 tahun. Istilah pentaksiran digunakan untuk menjalankan penilaian dalam bidang pendidikan di Malaysia. Untuk kajian ini, pengkaji telah mengadaptasi instrumen ini dan menambah beberapa perkara yang sesuai dengan pentaksiran kanak-kanak dalam pengajaran dan pembelajaran. Tiga perkara tersebut adalah Fokus, Standard Kandungan dan Standard Pembelajaran, iaitu seperti berikut:

Fokus didefinisikan sebagai bidang pembelajaran dalam instrumen yang ingin dikembangkan dalam diri kanak-kanak. Standard Kandungan didefinisikan sebagai pernyataan spesifik tentang perkara yang kanak-kanak patut tahu dan boleh lakukan dalam suatu tempoh tertentu. Standard Kandungan adalah huraian bagi Fokus. Standard Pembelajaran didefinisikan sebagai satu penetapan kriteria atau tahap pencapaian dalam bentuk tingkah laku yang akan diukur untuk memastikan penguasaan Standard Kandungan.

Item prestasi didefinisikan sebagai satu tugas yang meminta kanak-kanak memberi respon atau eviden, atau menyelesaikan masalah dalam bentuk tulisan, lisan, memanipulasi bahan dan ransangan yang diberikan, yakni perlakuan kanak-kanak

yang ditaksir secara langsung, bukan berbentuk persepsi (ya atau tidak) atau memilih jawapan (aneka pilihan). Semua item prestasi merujuk kepada fokus. Item prestasi adalah perkara yang akan ditaksir dan akan dikumpulkan sebagai hasil kerja sebagai portfolio.

**Pembelajaran Kanak-kanak.** Definisi konsep pembelajaran kanak-kanak adalah merujuk kepada perubahan kognitif dan tingkah laku kanak-kanak yang terhasil daripada pengalaman (Morrison, 2012). Dalam kajian ini, pembelajaran kanak-kanak yang merujuk kepada perkara yang ditaksir yang terkandung di dalam Kurikulum Standard Prasekolah Kebangsaan (tunjang) yang diselarikan dengan skop Kurikulum Kebangsaan dalam Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan (bidang pembelajaran) dan diadaptasi agar sesuai dengan kajian. Pembentukan domain-domain pembelajaran menggunakan Teori Kognitif Piaget dan Vygotsky dibincangkan di dalam Bab 2. Domain pembelajaran adalah Domain Kognitif, Domain Bahasa, Domain Kreativiti, Domain Fizikal, Domain Sosioemosi dan Domain Rohani.

**Perkembangan Kanak-kanak.** Definisi konsep perkembangan kanak-kanak adalah proses perubahan seseorang dalam tempoh tertentu. Perkembangan dipengaruhi oleh umur kronologi kanak-kanak, kadar kematangan dan pengalaman individu. Perkembangan boleh berubah sama ada kuantitatif seperti pertumbuhan fizikal atau kualitatif seperti perubahan emosi dan bahasa (Wortham, 2012). Dalam kajian ini, perkembangan kanak-kanak melalui beberapa peringkat untuk mencapai tahap maksimun. Pentaksiran perlu dilakukan untuk mengetahui kanak-kanak berada pada tahap mana kerana semakin sukar tugas yang diberikan semakin banyak bimbingan yang diperlukan. Melalui pentaksiran dapat mengetahui kanak-kanak dapat melakukan tugas dengan sempurna atau sebaliknya, mengenal pasti tugas tertentu hanya boleh dilakukan dengan bantuan atau sebaliknya dan mengesan tugas yang mudah atau sebaliknya. Penskoran akan diberi mengambilkira definisi operasi ini. Definisi

operasi perkembangan kanak-kanak dalam kajian ini merujuk kepada tiga perkara, iaitu, potensi perkembangan didefinisikan sebagai mencapai tahap tertinggi suatu tugas berjaya dilaksanakan oleh kanak-kanak dengan bantuan dan bimbingan diberikan. Perkembangan Sebenar didefinisikan sebagai mencapai tahap tertinggi suatu tugas berjaya dilaksanakan oleh kanak-kanak tanpa bantuan dan bimbingan. Perkembangan *above-level* didefinisikan sebagai mencapai tahap tertinggi suatu tugas berjaya dilakukan oleh kanak-kanak yang lebih muda dalam masa yang sama dilakukan oleh kanak-kanak yang lebih dewasa. Seseengah tugas mudah kepada kanak-kanak tapi tugas yang sama mencabar dan sukar bagi kanak-kanak yang lain walaupun dalam lingkungan umur yang berbeza.

**Pentaksiran Prestasi.** Definisi Konsep pentaksiran prestasi adalah pentaksiran yang memerlukan seseorang menunjukkan kemahiran dan pengetahuan yang dimilikinya sebagai maklum balas terhadap suatu tugas. Pentaksiran ini dilihat sebagai alternatif kepada penilaian aneka pilihan dan jawapan pendek yang hanya menumpukan kepada objektif pembelajaran (Russell & Airasian, 2012). Manakala dalam kajian ini, pentaksiran prestasi melibatkan proses pemerhatian dan pengujian item-item dengan meminta kanak-kanak memberi respon atau eviden, atau menyelesaikan masalah dalam bentuk bertulis, memanipulasi bahan dan ransangan yang diberikan, iaitu perlakuan yang dapat ditaksir secara langsung, bukan berbentuk persepsi atau memilih jawapan. Dalam kajian ini, instrumen yang telah diadaptasi ini disusun semula item mengikut domain perkembangan dan ditaksir. Maka item pentaksiran prestasi yang terdapat dalam instrumen adalah meliputi keseluruhan enam domain perkembangan.

### **Batasan Kajian**

Lokasi kajian ini dijalankan di taska-taska swasta dari dua daerah di negeri Selangor, iaitu Daerah Hulu Selangor (Rintis) dan Bandaraya Shah Alam (Selangor).

## **Rumusan**

Dalam bab 1, penyelidik melihat pembelajaran dan perkembangan kanak-kanak adalah pelbagai dan berbeza. Pendekatan untuk mentaksir prestasi mereka melalui instrumen yang dihasilkan mengikut budaya Malaysia dan menekankan pentaksiran autentik atau berasas prestasi. Pembinaan dan pengesahan instrumen pentaksiran prestasi, standard awal pembelajaran dan perkembangan awal kanak-kanak berdasarkan kerangka teoritikal dan konsep yang mengandungi empat faset yang akan dinilai serta mempunyai susunan prosedur yang sistematik dan tatacara statistik yang tepat untuk menghasilkan instrumen yang boleh digunakan sah dan dipercayai. Pentaksiran pada peringkat awal kanak-kanak sekarang amat penting dilakukan dengan mengambil kira model pengukuran yang diaplikasikan dan holistik supaya prestasi kanak-kanak pada masa sekarang dapat menggambarkan prestasi sebenar pada masa hadapan.

## Bab 2 Kajian Literatur

### Pengenalan

Bahagian ini menjelaskan tentang kajian literatur dan kajian berkaitan dengan pentaksiran berasas prestasi, pembentukan konsep dan definisi komponen faset pentaksiran berasas prestasi, penilaian pembelajaran dan perkembangan kanak-kanak, kepelbagaian latar belakang perkembangan kanak-kanak dari perspektif teori, pembinaan instrumen dan kajian luar negara berkaitan model pengukuran, pentaksiran perkembangan kanak-kanak di Malaysia dan teori pengujian moden.

### Definisi Standard Awal Pembelajaran dan Perkembangan Kanak-kanak

Terdapat pelbagai istilah yang berbeza digunakan untuk merujuk kepada standard awal. Perbezaan istilah bagi negara-negara tersebut adalah seperti yang ditunjukkan dalam Jadual 2.1. Istilah yang berbeza digunakan di beberapa buah negara untuk memerihalkan standard awal menghasilkan definisi yang agak berbagai-bagai bagi standard awal. Tidak ada definisi yang tetap bagi menyatakan maksud standard awal yang sebenarnya. Standard awal yang dibina berbeza berdasarkan keperluan dan keutamaan setiap negara (Junko & Cliff, 2008).

Jadual 2.1  
*Perbezaan Istilah Standard Awal Mengikut Negara*

Negara	Istilah yang digunakan
China	<i>Guidelines</i>
Thailand	<i>Competencies</i>
Cambodia	<i>School Readiness</i>

Walau bagaimanapun, apabila standard dihubungkan dengan pembelajaran dan perkembangan kanak-kanak, ia membawa maksud yang lebih khusus iaitu untuk menentukan apa yang kanak-kanak patut tahu dan boleh lakukan pada peringkat awal umur merentasi pelbagai domain perkembangan mengikut setiap negara. Kemudian, kefahaman tersebut dapat menyediakan asas bagi sesebuah negara untuk mereka

bentuk kurikulum, untuk menyemak semula program pengasuhan atau latihan, untuk mengkaji semula usaha pendidikan dan sokongan keibubapaan yang semua akan menyumbang kepada kanak-kanak mencapai potensi perkembangan diri mereka.

Menurut laporan *National Institute for Child and Family Development* pada tahun 2003, walaupun terdapat pelbagai alat penilaian perkembangan kanak-kanak di negara-negara ASEAN, namun masih kekurangan alat penilaian yang sesuai untuk perkembangan psikososial yang mengukur perkembangan kanak-kanak secara holistik. Kebiasaannya, negara-negara ini akan menggunakan alat penilaian yang dibangunkan oleh negara barat secara adaptasi atau sebaliknya. Pada awal tahun 2005, UNICEF telah memulakan kerjasama dengan tujuh buah negara ASEAN untuk memulakan proses membina standard awal pembelajaran dan perkembangan kanak-kanak mengikut acuan dan budaya negara tersebut dan sehingga kini hampir 40 buah negara telah memulakan proses membina standard awal pembelajaran dan perkembangan kanak-kanak dengan bantuan UNICEF.

Konsep standard awal pembelajaran dan perkembangan kanak-kanak yang diperkenalkan seperti di atas menjadikan pencapaian kanak-kanak sebagai penanda aras. Definisi standard ini menggambarkan pengetahuan sedia ada kanak-kanak dan kebolehan yang dimiliki semasa peringkat awal perkembangan. Standard membolehkan maklumat yang akan dikumpul melalui pemerhatian langsung terhadap kanak-kanak dan sebagai asas perbandingan dalam mengukur dan menilai kapasiti, kualiti, nilai, atau kuantiti (Kagan & Britto, 2005).

Standard awal pembelajaran dan perkembangan ini dibenarkan setiap negara untuk menentukan kerangka kerja domain sendiri dengan menggunakan istilah mereka sendiri dan perspektif kebangsaan yang akan menentukan domain pembelajaran dan perkembangan kanak-kanak secara holistik. Setiap standard memberikan maklumat mengenai tahap tingkah laku, pencapaian bahasa dan kognitif untuk kanak-kanak pada



masa yang diberikan. Pendekatan ini diasaskan kepada bukti penyelidikan mengenai perkembangan kanak-kanak dan mengambil kira perbezaan budaya, bahasa, dan sosioekonomi pertimbangan, serta keperluan kanak-kanak khas (Kagan & Britto, 2005a).

Junko & Cliff (2008) menghuraikan objektif standard awal pembelajaran dan perkembangan kanak-kanak bagi kebanyakan Negara yang terlibat adalah untuk membina kerangka yang komprehensif untuk awal kanak-kanak yang relevan dengan konteks kebangsaan. Proses pembinaan ini perlu memenuhi perkembangan kanak-kanak secara menyeluruh atau holistik secara semulajadi yang sesuai dengan budaya, norma sosial dan keluarga. Kepentingan terhadap pembinaan standard awal ini bukan sahaja untuk melihat kualiti standard tersebut, malah memberi kesan yang positif kepada kumpulan sasaran yang akan menggunakan alat instrumen ini dan kepada siapa ia ditadbir.

Menghubungkan standard awal dan kurikulum juga merupakan objektif yang relevan bagi beberapa buah negara, iaitu apabila konsep pengajaran jelas, mempunyai jangkaan terhadap perkara yang dipelajari oleh kanak-kanak dan bagaimana kanak-kanak menunjukkan apa yang telah mereka pelajari, maka kurikulum kebangsaan akan mengambil kira hasil pembelajaran dan perkembangan kanak-kanak. Seterusnya membantu pihak yang menjadi pengamal dalam bidang awal kanak-kanak meletakkan fokus utama terhadap pelaksanaan pengajaran dengan lebih sistematik (Scott-Little, Kagan & Frelow, 2003).

### **Model Standard Awal Pembelajaran dan Perkembangan**

Pada tahun 2002, UNICEF dalam kerjasama dengan Universiti Colombia dan Universiti Yale melancarkan projek bertajuk “ *Going Global with Early Learning and Development Standard*” untuk mengatasi masalah ketidaksesuaian instrumen untuk menilai dan memantau perkembangan awal kanak-kanak. Pada pertengahan tahun

2005, negara seperti Cambodia, China, Fiji, Mongolia, Filipina, Thailand dan Vietnam telah menunjukkan minat untuk mendefinisikan dan menentukan satu set standard awal pembelajaran dan perkembangan pada peringkat kanak-kanak. Usaha di ketujuh-tujuh negara ini mula berkembang selaras dengan keperluan di negara masing-masing dan indikator khusus diperlukan untuk memastikan bahawa standard awal pembelajaran dan perkembangan kanak-kanak bagi menjalankan penilaian dipenuhi. Di Malaysia, perkembangan ini bermula di sekitar tahun 2010. Pihak Kementerian Pelajaran Malaysia menerusi Bahagian Pembangunan Kurikulum telah memulakan usaha membangunkan standard ini menerusi dana UNICEF.

Standard awal pembelajaran dan perkembangan adalah model asas untuk membina alat yang membantu meningkatkan amalan pengajaran di antara pendidik awal kanak-kanak. Mempunyai kefahaman kepada domain, standard dan bagaimana proses pemerhatian dijalankan, adalah merupakan asas untuk kandungan pedagogi pengajaran. Sementara itu, membina alat yang sesuai untuk membezakan tahap perkembangan kanak-kanak merentasi semua domain, membolehkan pendidik awal kanak-kanak untuk merancang sesi dan persekitaran pembelajaran yang lebih sesuai. Vietnam merupakan contoh negara yang membuat perancangan membina standard untuk guru-guru berdasarkan hasil pembelajaran daripada standard awal pembelajaran dan perkembangan ini.

Filipina menggunakan standard awal pembelajaran dan perkembangan untuk menstabilkan sistem bagi *Early Childhood Development (ECD)* yang sedia ada dan memurnikan ia dengan semua standard dan alat penilaian atau pengukuran seperti senarai semak guru, standard guru dan standard pekerja. Di Mongolia pula, standard awal pembelajaran dan perkembangan dibina untuk menghubungkan polisi kebangsaan pendidikan awal kanak-kanak dengan peningkatan kualiti untuk menyokong perkembangan holistik dan keperluan kanak-kanak. Pelaksanaan yang

konsisten semasa peringkat awal kanak-kanak hingga pra sekolah membantu dalam melihat kesan jangka masa panjang. Seperti di Cambodia dan Mongolia juga, keselarian pelaksanaan peringkat awal hingga pra-sekolah ditekankan untuk memastikan standard di peringkat seterusnya juga memberi kesan kepada perkembangan holistik kanak-kanak.

China dan Fiji mengambil pendekatan yang sedikit berbeza semasa membina standard ini, iaitu bertujuan untuk menjadikan model ini berupaya meningkatkan dasar kebangsaan terhadap perkembangan kesediaan kanak-kanak dan melakukan pemantauan. Cambodia dan Laos menjalankan penyelidikan untuk melihat keberkesanan pendidikan awal kanak-kanak di negara mereka melalui kajian rintis pada tahun 2005 menggunakan indikator tertentu dalam standard awal pembelajaran dan perkembangan yang dibina. Hasil kajian tersebut menunjukkan penyertaan dalam program pendidikan awal kanak-kanak sama ada berbentuk program komuniti, program di rumah atau di sekolah, menunjukkan hubungan positif yang signifikan terhadap perkembangan awal kanak-kanak (Rao & Person, 2007).

Pentaksiran adalah satu proses yang sistematik untuk mengumpul maklumat tentang perkembangan murid bagi membantu guru membuat penilaian terhadap mereka dengan tujuan mengenal pasti tindakan susulan yang sesuai untuk membantu murid mencapai objektif pengajaran dan pembelajaran. Proses dalam pentaksiran melibatkan aktiviti mengumpul, menganalisis dan menginterpretasi maklumat yang diperoleh.

Pentaksiran adalah satu proses yang berterusan dan dijadikan sebahagian daripada pengajaran dan pembelajaran. Dalam pentaksiran, guru boleh menggunakan pelbagai kaedah untuk mendapatkan maklumat. Hasil pentaksiran terhadap murid membolehkan guru mengambil tindakan susulan yang berkesan dan seterusnya membantu guru memperkembangkan potensi murid sepenuhnya dalam domain

kognitif, afektif dan psikomotor sebagai persediaan mereka untuk ke prasekolah dan seterusnya ke sekolah rendah.

Pentaksiran kanak-kanak di Malaysia umumnya bukan bertujuan melabel dan membandingkan mereka antara satu sama lain, tetapi untuk meningkatkan pembelajaran murid berkenaan. Istilah yang sering digunakan untuk pentaksiran seperti ini ialah Pentaksiran Autentik atau Pentaksiran secara Formatif. Pentaksiran kanak-kanak lebih menekankan kepada usaha ke arah penambahbaikan serta perkembangan dan bukan pencapaian semata-mata. Maklum balas yang diberi semasa atau selepas sesuatu aktiviti pentaksiran merupakan faktor penting dalam menjayakan proses pentaksiran. Pertimbangan harus diberi ke atas keperluan kanak-kanak, perasaan kanak-kanak, psikologi kanak-kanak, teori perkembangan kanak-kanak semasa merancang proses pentaksiran.

Hasil pentaksiran yang telah diperolehi akan merujuk kepada Standard Prestasi. Standard Prestasi kanak-kanak di Malaysia adalah merupakan satu set kriteria umum yang menunjukkan tahap-tahap prestasi yang perlu murid pamerkan sebagai tanda bahawa, sesuatu perkara telah dikuasai oleh murid. Ianya memberi definisi kepada “kejayaan” dan bagaimanakah hasil kerja dan tingkah laku yang dikatakan cukup baik (*indicator of success*) atau sudah mencapai objektif pembelajaran. Tujuan diadakan standard prestasi ialah untuk mengenal pasti di manakah tahap penguasaan pengetahuan, kemahiran atau pengamalan nilai seseorang murid itu. Standard prestasi memberi gambaran jelas atau panduan kepada guru tentang bagaimana hendak memberi skor kepada hasil yang dipersembahkan oleh murid. Standard Prestasi juga dirujuk semasa guru memberi maklum balas dan membuat pelaporan tentang murid.

Standard Prestasi memperihalkan kualiti penguasaan sesuatu perkara atau hasil kerja mengikut hierarki. Standard Prestasi tidak menjelaskan secara spesifik tentang

teknik atau kaedah untuk mencapai hasil yang dikehendaki. Standard Prestasi bukan saja patut diketahui oleh pentaksir (guru), malah ibu bapa dan murid sendiri.

Standard Prestasi dinyatakan dalam bentuk kriteria dan skor. Kriteria merupakan pernyataan yang menjelaskan tentang kualiti sesuatu kerja atau perkara yang dikuasai. Kriteria yang ditulis mesti mudah, ringkas, tidak mengelirukan dengan mengambil kira perbezaan antara individu. Ianya menjadi rujukan bagaimana prestasi murid ditaksir dan diberikan skor. Pernyataan kriteria yang ditulis pada setiap tahap penguasaan mestilah dapat memperihalkan setakat mana kualiti kerja yang ditunjukkan oleh murid. Kriteria yang khusus dibina bagi setiap konstruk yang dinilai dalam pentaksiran kanak-kanak. Bagi memastikan kesamaan kualiti antara kesemua kriteria khusus, satu standard dan kriteria teras (*core*) digunakan sebagai panduan.

Dalam konteks pentaksiran di Malaysia, Skor bermaksud apa yang digunakan untuk menandakan tahap penguasaan murid. Sistem pentaksiran untuk kanak-kanak telah dibangunkan pada 2010 (Bahagian Pembangunan Kurikulum, 2010) dan dilaksanakan pada peringkat prasekolah selari dengan Kurikulum Standard Prasekolah Kebangsaan atau KSPK. Di prasekolah, skor yang diberi bukan sekadar berbentuk gred (A,B, C,...) atau markah (1,2,3,...) yang hanya boleh dijumlah dan dipertakan, seperti yang digunakan dalam ujian pencapaian di peperiksaan awam. Skor dalam bentuk markah dan gred semata-mata tidak boleh memberi maklumat tentang apa yang murid boleh lakukan. Untuk menjadikan pentaksiran lebih bermakna di prasekolah, skor mestilah dapat menjelaskan di tahap manakah murid berada berbanding standard prestasi yang hendak dicapai. Contoh skor yang boleh digunakan dalam pentaksiran prasekolah ialah seperti dalam Jadual 2.2 berikut:

## Jadual 2.2

### *Jadual Standard Prestasi*

Skor (Tahap Penguasaan)	Simbol	Kriteria
Menguasai	M	Boleh melakukan aktiviti dengan betul sehingga selesai.
Sedang Berkembang	SB	Memerlukan bimbingan dari semasa ke semasa apabila melakukannya sehingga selesai
Perlu Banyak Bimbingan	PB	Sukar memahami arahan untuk menjalankan aktiviti dan sentiasa memerlukan bimbingan

Dengan adanya Standard Prestasi, guru dapat memberi skor dan maklum balas yang tepat kepada murid tentang prestasi mereka. Sehubungan itu, guru boleh mengatur strategi pengajaran atau intervensi untuk membantu murid memperbaiki pembelajaran mereka bagi menguasai tahap yang tertinggi. Daripada maklum balas guru, murid pula mendapat gambaran yang jelas tentang sasaran pembelajaran, di mana kedudukan mereka dan apa lagi yang mereka perlu perbaiki untuk maju ke arah sasaran berkenaan.

### **Domain Standard Awal Pembelajaran dan Perkembangan Kanak-kanak yang Diukur**

Dari segi rekabentuk domain, konsep yang digunakan diselarikan dengan amalan negara masing-masing. Setiap negara menggunakan konsep 'holistik' dan bermula dengan 3 hingga 7 domain. Negara Mongolia bermula dengan 7 domain kemudian mengurangkan kepada 3 domain. Bagi kandungan domain, pembangunan *Early Learning Development Standard* (ELDS) meliputi perkembangan motor dan kesihatan fizikal, perkembangan sosioemosi, perkembangan kognitif dan maklumat, perkembangan bahasa, kreativiti dan seni.

Standard Awal Pembelajaran Perkembangan (Malaysia) dibangunkan oleh kerjasama panel pakar seluruh negara melalui Bahagian Pembangunan Kurikulum, Kementerian Pendidikan Malaysia (2003). Penyelidik memilih untuk membangunkan enam 6 domain iaitu Kognitif, Bahasa, Kreativiti, Fizikal, Sosioemosi dan Rohani berdasarkan standard ini. Dalam kajian ini, instrumen yang telah dibina, ditadbir

semula untuk mencari nilai kesahan dan kebolehpercayaan instrumen serta, mengenal pasti empat faset utama yang ditetapkan dalam objektif kajian seterusnya disokong oleh kajian literatur bagi setiap domain.

### **Perkembangan Kognitif**

*Pembelajaran.* Pembelajaran secara analogi adalah satu lagi pendekatan penting dalam pengajaran dan pembelajaran yang terdapat pada awal perkembangan kanak-kanak. Analogi melibatkan persamaan antara satu keadaan dengan satu lagi yang lain, atau antara satu masalah dengan masalah yang lain. Persamaan ini kemudian menjadi asas untuk memerlukan penyelesaian analogi. Chen, Sanchez & Campbell (1997) dalam kajiannya, mengatakan keupayaan kanak-kanak untuk belajar dengan analogi boleh diuji menggunakan prosedur penyelesaian masalah yang mudah. Beliau menunjukkan bahawa kanak-kanak berusia 13 bulan boleh menggunakan analogi untuk menyelesaikan masalah ini. Kanak-kanak boleh menyelesaikan analogi yang sama dalam situasi yang lebih rumit (Brown, 1990).

Pembelajaran sebab dan akibat juga merupakan pembelajaran di peringkat awal. Pembelajaran ini disebut pembelajaran berasaskan huraian. Baillargeon, Li, Ng & Yuan (2009) dalam kajiannya telah berhujah bahawa bayi berhadapan dengan masalah yang sama untuk belajar tentang dunia fizikal. Sebagai contoh, mereka melihat pelbagai pergerakan satu fenomena tertentu, seperti objek yang jatuh, dan perlu mengetahui apa yang menyebabkan ia jatuh.

Neurosains kognitif menyatakan semua bentuk pembelajaran penting bagi kognitif manusia dengan itu hadir dalam pelbagai cara. Pembelajaran secara analogi dan pembelajaran sebab dan akibat menyokong perkembangan kognitif. Perkembangan neurosains kognitif mendedahkan bagaimana kekuatan mekanisme pembelajaran adalah, menjadikan pembelajaran kanak-kanak lebih cepat mengenai rangsangan sosial seperti reaksi muka (Farroni, Csibra, Simion & Johnson, 2002),

fizikal seperti tindakan menggenggam (Tai, Scherfler, Brooks, Sawamoto & Castiello, 2004), dan bahasa (Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman & Dehaene, 2006).

***Pembinaan pengetahuan.*** Kanak-kanak membina pengetahuan tentang sebab dan akibat daripada maklumat persepsi (Leslie, 1994). Penemuan saintifik biasanya difahami sebagai jenis pemikiran yang memerlukan penyelarasan dan pembezaan teori dan bukti, dan penilaian hipotesis (Kuhn, 1989). Kanak-kanak memerlukan masa untuk membina pengetahuan penaakulan saintifik yang baru kerana ia menambah ke dalam pengetahuan sedia ada mereka. Perkara ini disebabkan mereka perlu menyimpan beberapa perkara baru dalam fikiran sekaligus (Kuhn, 1995). Kajian Goswami (2014) menunjukkan perkembangan ini, terutama dalam membina pengetahuan sains akan mempengaruhi mereka apabila dewasa. Iaitu pemikiran dan hujah mereka boleh menjadi lebih berkesan kerana pengetahuan yang baik.

Kanak-kanak mengetahui, objek yang bergerak sendiri adalah agen bernyawa, dan pergerakan mereka tidak boleh diramal tetapi adalah disebabkan oleh bahagian dalaman mereka sendiri (Gelman & Opfer, 2010). Mengetahui ciri-ciri fizikal seperti lebih tinggi atau gemuk, boleh menukar warna atau bentuk, contohnya ulat kepada rama-rama, dan bahasa membantukanak-kanak untuk memberi tumpuan kepada persamaan struktur dan fungsi, kerana konsistensi pelabelan objek itu contohnya 'burung' untuk burung unta, 'haiwan' untuk kucing dan lembu.

***Memori.*** Memori terdiri daripada pelbagai aspek dalam perkembangan kognitif. Antara yang paling utama adalah memori semantik (generik, fakta pengetahuan kita tentang dunia), memori episod (keupayaan kita secara sedar untuk mendapatkan semula kejadian dari masa lalu), tersirat atau memori prosedur (seperti tabiat dan kemahiran), dan ingatan kerja. Kenangan yang boleh dibawa secara sedar dan sengaja di fikiran (ingatan semantik dan episod) dipanggil perisytiharan,



sedangkan pengetahuan yang biasanya ditentukan oleh perubahan dalam prestasi (misalnya menunggang basikal) dipanggil memori tersirat. Sebagai contoh, dalam kajian Brown & Scott (1971) menunjukkan kanak-kanak berumur 3 hingga 5 tahun mempunyai memori 98 peratus kepada satu set gambar.

Perkembangan memori terletak dalam aktiviti sosial dan kognitif, oleh itu struktur pengetahuan kanak-kanak bergantung kepada pengalaman mereka. Kajian Bauer (2010) menunjukkan walaupun berusia 18 bulan, kanak-kanak pada usia ini mengekalkan ingatan yang memaparkan perkara yang pernah dilihat dan berlaku. Nelson (1986) menunjukkan bahawa kanak-kanak yang lebih kecil menumpukan perhatian kepada rutin mengingat, sebagai rutin menjadikan dunia ini tempat yang boleh diramal. Walau bagaimanapun, kanak-kanak kecil juga mengingati peristiwa yang tersendiri. Dalam kajian Fivush dan Hamond (1990) melaporkan bahawa kanak-kanak berusia 4 tahun yang teringat bahawa, ketika beliau berumur 2 setengah tahun 'Saya makan ikan saya dan ibu saya telah membuang sisanya. Walau bagaimanapun, kanak-kanak kecil jarang mencipta kenangan yang belum berlaku (Gilstrap & Ceci 2005).

**Bahasa.** Bahasa memainkan peranan penting dalam perkembangan kognitif. Bahasa membantu perkembangan teori, fikiran, perkembangan memori dan adalah asas ingatan kerja. Ia juga perkara penting dalam teori Vygotsky yang berkaitan dengan perkembangan kognitif. Kanak-kanak menggunakan kebolehan yang sama untuk memperoleh fonologi (berasaskan bunyi) aspek bahasa yang mereka gunakan untuk memperoleh pengetahuan tentang dunia fizikal dan psikologi, iaitu pembelajaran perhubungan, dan membuat analogi.

Penggunaan perbendaharaan kata sebagai fungsi utama bahasa adalah komunikasi, dan kata-kata adalah sebahagian daripada pengalaman yang bermakna. Satu kajian menunjukkan bahawa kanak-kanak mendengar anggaran 5000-7000

ucapan-ucapan sehari, dengan kira-kira satu pertiga daripada ucapan-ucapan ini menjadi soalan (Cameron-Faulkner, Lieven & Tomasello, 2003). Dalam kajian di Amerika Syarikat, Hart & Risley (1995) menganggarkan bahawa kanak-kanak dari status sosio-ekonomi keluarga yang berada secara puratanya mendengar 487 ucapan sejam, berbanding 178 ucapan sahaja dalam sejam bagi kanak-kanak daripada keluarga yang tinggal di rumah kebajikan. Oleh itu dengan masa yang mereka telah jangkau iaitu berumur 4 tahun, kanak-kanak telah terdedah kepada kira-kira 44 juta ucapan-ucapan, berbanding 12 juta ucapan-ucapan untuk kanak-kanak berstatus sosio-ekonomi rendah.

Pembelajaran perkataan juga penting untuk perkembangan kognitif kerana ia adalah peringkat simbolik. Perkataan adalah simbol kerana mereka merujuk kepada objek atau peristiwa, tetapi mereka tidak menyebut objek atau peristiwa itu sendiri. Simbol membolehkan kanak-kanak untuk mengenali diri mereka daripada keadaan tertentu. Gerak isyarat juga simbolik (contohnya melambai 'selamat tinggal'). Gerak Isyarat mendahului dalam perkembangan pengeluaran bahasa, menyediakan 'jambatan kognitif' antara kefahaman dan maklum balas (Voltera & Erting, 1990). Tindakan digunakan untuk menyatakan makna. Walaupun kemudian dalam perkembangan kognitif, isyarat boleh memberikan maklumat penting tentang apa yang anak itu memahaminya dalam domain perkembangan kognitif yang diberikan. Kajian oleh Goldin-Meadow & Wagner (2005) mengenai gerak isyarat tidak sedar yang ditunjukkan oleh kanak-kanak telah diperhatikan oleh guru-guru mereka, yang mengubah input pengajaran mereka dengan maklum balas sewajarnya. Terdapat 'ketidakpadanan' isyarat-ucapan sering dijumpai ketika kanak-kanak sedang melakukan tugas tertentu berkaitan perkembangan kognitif.

Penggunaan perkataan (perkembangan perbendaharaan kata) adalah respon pada perkembangan awal kanak-kanak. Menggunakan senarai semak bahasa kanak-

kanak (kini diterjemahkan ke dalam lebih daripada 20 bahasa), kajian Fenson, Dale, Reznick, Bates, Thal & Pethick (1994) menunjukkan bahawa bahasa yang dituturkan mengandungi saiz perbendaharaan kata adalah sebanyak 55 perkataan pada 16 bulan, 225 perkataan pada 23 bulan, 573 perkataan pada 30 bulan, dan 6000 perkataan pada umur 6 tahun. Kefahaman perbendaharaan kata pada usia 6 tahun adalah sekitar 14,000 perkataan (Dollaghan, 1994). Sebagai contoh, pada umur 2 tahun, kepelbagaian dalam pengeluaran perkataan adalah dari 0 kata-kata kepada lebih daripada 500 patah perkataan. Kajian Fenson juga menunjukkan bahawa dalam pemerolehan perbendaharaan kata pada umur 18 bulan berlaku bagi kebanyakan kanak-kanak. Maksud perkataan melalui pengenalan nama-nama perkataan adalah penting secara teori, kerana ia menunjukkan pencapaian kognitif secara teratur (Bloom, 1973). Pencapaian pada umur 18 bulan kelihatan sesuai dan selari dengan pandangan teori Piaget yang membawa pemahaman yang simbolik kepada 'konsep objek'. Walau bagaimanapun, bayi seawal usia 4 bulan seolah-olah telah memahami bahawa perkataan boleh mempunyai nama atau maksud. Mereka sudah mengenali nama-nama mereka sendiri, dan perkataan untuk 'mummy' (Mandel, Jusczyk & Pisoni, 1995). Perkembangan perkataan baru adalah amat pesat, dengan kira-kira 10 perkataan baru yang diperolehi setiap hari pada usia 2 tahun.

***Pretend play dan imaginasi.*** *Pretend Play* adalah tindakan berpura-pura dalam bermain. Ia adalah kebolehan terawal kanak-kanak untuk memberi sifat atau ciri tertentu yang menghubungkan kognitif mereka dengan pengetahuan baru. Tindakan berpura-pura juga satu cara untuk melepaskan diri dari keadaan yang tertentu. Kajian Leslie (1987) menunjukkan bahawa tindakan berpura-pura seperti menjadikan pisang sebagai telefon, kanak-kanak mesti menukarkan ciri utama pisang (berwarna kuning dengan tekstur tertentu dan mempunyai bau tertentu) kepada tindakan berpura-pura iaitu melakonkan watak penerima telefon. Mewakili sesuatu objek secara langsung

adalah penting untuk perkembangan kognitif kerana kanak-kanak menganggap ia adalah situasi sebenar. Ia memerlukan kerjasama semasa iaitu dengan mengambil contoh telefon tadi, pemanggil dan penerima telefon bersama-sama mewakili situasi sebenar dengan objek tertentu. Maka, *pretend play* ini, ia merupakan permulaan kebolehan untuk memahami kognitif kanak-kanak itu sendiri, serta memahami bahawa kanak-kanak itu berfikir objek sebagai satu entiti.

“*Pretend play*” akan berkembang pada umur 2 tahun, iaitu bermula dengan sifat berpura-pura yang pada biasanya terikat kepada tindakan yang orang dewasa lakukan pada objek sebenar. Contohnya, kanak-kanak boleh 'minum' pada umur 12 bulan daripada cawan kosong. Selain itu, kanak-kanak yang berusia 2 tahun berpura-pura menjadikan sebatang kayu sebagai kuda. Aktiviti ini biasanya dilakukan dengan orang lain. Kanak-kanak menunjukkan tindakan yang lebih menonjol apabila mereka meniru dan berpura-pura bersama orang lain, dan sebaiknya orang dewasa perlu turut bermain bermain agar membantu lagi perkembangan kefahaman simbolik (Bigelow, MacLean & Proctor, 2004). Bahasa juga memainkan peranan penting, sebagai rakan kongsi sosial, kanak-kanak boleh menggunakan bahasa untuk membantu menjelaskan keadaan *pretend play* tersebut, atau boleh menggunakan "suara ganjil" untuk memberi isyarat bahawa keadaan adalah berpura-pura.

*Pretend play* juga dikaitkan dengan perkembangan ‘*Theory of Mind*’ (TOM) oleh Vygotsky. Walau bagaimanapun, kepelbagaian rakan-rakan sosial yang berbeza mewujudkan pelbagai jenis permainan. *Pretend play* dengan adik-beradik atau kawan-kawan tidak sama dengan bermain berpura-pura dengan ibu, dan lebih cenderung untuk menjadi lebih sosial, kerana kanak-kanak lain biasanya membawa watak tertentu dalam drama itu sendiri. Kajian Jenkins & Astington (1996) menunjukkan bahawa kanak-kanak yang mempunyai adik-beradik menunjukkan perkembangan TOM lebih awal daripada kanak-kanak tanpa adik-beradik, dan yang mempunyai adik-beradik

mempunyai kesan yang lebih kukuh. Salah satu sebab *pretend play* bersama adik-beradik dan rakan-rakan adalah untuk membantu meningkatkan pemahaman psikologi terhadap kebaikan *pretend play* yang menggalakkan imaginasi yang tinggi dan kerjasama yang baik.

Apabila kanak-kanak semakin membesar, masa juga kurang diperuntukan dalam aktiviti permainan yang sebenar, dan lebih banyak masa dihabiskan dalam memikirkan peranan masing-masing (Lillard, 2002). Keadaan ini menunjukkan kemahiran minda meningkat untuk membaca. Hughes dan Dunn (1998) menunjukkan dalam kajiannya, kanak-kanak berusia 4 tahun bahawa keadaan mental kanak-kanak dan rakan-rakan di taska adalah signifikan dengan prestasi masing-masing. Peranan imaginasi dalam perkembangan kognitif sangat penting kerana memberi tumpuan kepada pemahaman yang lebih mendalam terhadap fikiran (perkembangan kognitif sosial). Vygotsky (1978) menyatakan bahawa imaginasi itu mewakili satu bentuk khusus manusia dalam aktiviti kognitif. Menurut teori beliau, fungsi pembangunan pusat *pretend play* adalah kanak-kanak perlu bertindak terhadap rangsangan dan ikut 'peraturan permainan. Ini dianggap untuk membantu mereka untuk mendapatkan kawalan ke atas pemikiran dan tindakan mereka. Rakan kanak-kanak yang bermain bersama memberi fungsi kawal selia yang penting juga. Kajian Karpov (2005) melaporkan kanak-kanak berumur 3 hingga 7 tahun berada dalam satu permainan perlu untuk berdiri dan tidak bergerak untuk seberapa lama yang boleh dalam satu permainan. Terdapat kanak-kanak yang terus main. Apabila kanak-kanak terpaksa berdiri kaku sahaja di dalam bilik tanpa konteks bermain, mereka kurang aktif berbanding dengan yang terlibat terus dengan konteks permainan. Walau bagaimanapun, apabila mereka terpaksa senyap dalam bilik yang penuh dengan rakan sepermainannya, mereka yang paling baik perkembangan kognitifnya. Rakan sepermainannya sedang memantau pergerakan, membantunya untuk berdiam diri

untuk lebih lama. Vygotsky menganggap bermain sebagai faktor utama dalam perkembangan kognitif.

**Perkembangan Bahasa.** Bahasa adalah satu bentuk komunikasi, sama ada yang dituturkan, bertulis, atau ditandatangani, yang berasaskan sistem simbol. Semua bahasa manusia adalah generatif. Generatif adalah keupayaan untuk menghasilkan nombor tanpa henti ayat yang bermakna menggunakan satu hingga dalam kata-kata dan peraturan sehingga menghasilkan bahasa yang sangat kreatif. Semua bahasa manusia juga mengikut peraturan organisasi fonologi, morfologi, sintaksis, semantik, dan pragmatik. bahasa yang dituturkan terdiri daripada bunyi-bunyi asas, atau fonem. Contoh fonem dalam bahasa Inggeris adalah / k /, bunyi diwakili oleh huruf k dalam ski perkataan, huruf c dalam perkataan kucing, dan huruf ch dalam Krismas. Fonologi adalah sistem bunyi bahasa. Peraturan fonologi membolehkan beberapa urutan bunyi berlaku (seperti sp, ba, atau ar) dan melarang orang lain (seperti zx atau qp). Untuk mengetahui fonologi bahasa, kanak-kanak mesti belajar inventori bunyi dan urutan penting untuk membaca (Oller, 2000).

Chomsky (1957) berhujah bahawa manusia belajar bahasa pada satu masa tertentu dan dalam cara yang tertentu. Bukti kuat untuk asas biologi bahasa adalah bahawa kanak-kanak di seluruh dunia memperoleh pencapaian bahasa pada masa yang sama dengan perkembangan dan dalam kira-kira usia yang sama, walaupun variasi yang luas dalam input bahasa yang mereka terima. Sebagai contoh, dalam sesetengah budaya orang dewasa tidak pernah bercakap kepada bayi mereka di bawah umur satu tahun, tetapi bayi-bayi ini masih memperoleh bahasa. Kanak-kanak berbeza dari segi pengambilalihan bahasa dalam cara yang tidak dapat dijelaskan oleh input alam sekitar sahaja.

Kajian mendapati bahawa mereka kadang-kadang tersenyum dan memuji anak-anak mereka untuk ayat yang mereka suka, tetapi mereka juga diperkukuhkan ayat yang bersama-sama mematuhi tatabahasa. Dari pemerhatian ini, disimpulkan bahawa proses yang beroperasi di dalam minda kanak-kanak itu telah mengatasi input alam sekitar dan luaran. Walau bagaimanapun, kanak-kanak tidak belajar bahasa dalam persekitaran sosial. Kepelbagaian yang berlaku dalam perkembangan bahasa apabila ibu bapa kanak-kanak menerapkan input bahasa yang berbeza dan tahu bahawa persekitaran memainkan peranan penting dalam perkembangan bahasa mereka, terutamanya dalam pemerolehan perbendaharaan kata. Malah sebelum mereka pergi ke sekolah, kebanyakan minda kanak-kanak telah penuh dengan bahasa.

Perkembangan bahasa bukan sekadar mengatakan perkara yang betul dan meniru pertuturan bahasa. Kanak-kanak mendapat manfaat apabila ibu bapa dan guru-guru mereka secara aktif melibatkan diri mereka dalam perbualan, bertanya kepada mereka soalan, dan menekankan interaktif bukannya bahasa arahan. Kanak-kanak berusia tiga tahun yang hidup dalam kemiskinan menunjukkan pengurangan perbendaharaan kata berbanding rakan-rakan mereka dalam keluarga berpendapatan sederhana, dan bahawa pengurangan ini kekal apabila mereka memasuki sekolah semasa enam tahun. Tingkah laku lisan dan kesannya terhadap bahasa telah dikaji. Didapati bahawa lebih daripada dua pertiga daripada kanak tadika tidak berkomunikasi dalam jangka tertentu.

Perkembangan bahasa kanak-kanak melalui beberapa peristiwa penting (Bloom, 1998). *Babbling* bermula pada kira-kira 3 hingga 6 bulan. Bayi biasanya menyebut perkataan pertama mereka pada kira-kira 10 hingga 13 bulan. Sebanyak 18 hingga 24 bulan, bayi biasanya telah mula rentetan dua perkataan bersama-sama. Pada peringkat dua perkataan ini, mereka dengan cepat memahami kepentingan bahasa

dalam komunikasi, mewujudkan frasa seperti "Buku di sana," "gula-gula saya," "Mama berjalan kaki," dan "Beri Papa."

Oleh kerana kanak-kanak dapat bertutur dengan 2 suku kata perkataan, kanak-kanak jelas menunjukkan bahawa mereka tahu beberapa peraturan morfologi (peraturan-peraturan bahasa yang memberitahu bagaimana bunyi mesti digabungkan). Kanak-kanak mula menggunakan bentuk jamak dan kata nama; meletakkan pengakhiran yang sesuai pada kata kerja (apabila subjek adalah orang ketiga tunggal, -ed untuk masa lalu tegang); penggunaan kata sendi (di dalam dan di), artikel (dan), dan pelbagai bentuk kata kerja seperti ( "Saya telah pergi ke kedai"). Bukti terbaik adalah kanak-kanak mengikut peraturan morfologi dan bukan menghafal perkataan.

**Perkembangan Fizikal.** Perkembangan motor kasar dan halus merupakan kemahiran penting dalam perkembangan fizikal. Motor kasar dan kemahiran motor halus berkembang semasa zaman kanak-kanak. Motor kasar kemahiran melibatkan aktiviti besar-otot, seperti berlari dan bermain bola keranjang. Kemahiran motor halus melibatkan pergerakan halus seperti ketangkasan jari yang diperlukan untuk menulis dan melukis. Guru taska dan tadika perlu melaksanakan aktiviti-aktiviti yang sesuai dengan perkembangan mereka yang melibatkan kemahiran motor kasar. Termasuk latihan yang melibatkan pergerakan asas, kecergasan harian dan aktiviti motor. Berjalan adalah satu contoh usaha bagi menggalakkan kemahiran pergerakan asas. Aktiviti kecergasan setiap hari boleh dilaksanakan dengan diiringi oleh muzik. Menggabungkan kecergasan dengan pergerakan kreatif, muzik, dan imaginasi kanak-kanak adalah strategi yang baik. Kanak-kanak melakukan pergerakan seperti ular, kucing, kanggaru, dan kapal terbang.

Guru boleh meminta kanak-kanak untuk meniru pergerakan mereka. Pembangunan kemahiran motor halus kanak-kanak membolehkan mereka untuk lebih meningkat dalam aspek kawalan dan kemahiran. Kanak-kanak mempunyai persepsi



yang besar terhadap seni iaitu bagaimana mereka melihat ruang dan jarak, dan bagaimana mereka membuat corak dan bentuk. Guru boleh memberi kanak-kanak ruang kerja di mana mereka tidak perlu bimbang tentang kekotoran dan barang yang berselerak.

**Perkembangan Kreativiti.** Kreativiti adalah satu fenomena pelbagai dimensi dan kompleks. Ia adalah sukar untuk diramal dan salah satu konsep psikologi yang paling sukar untuk ditentukan (Kousoulas, 2010; Kurtzberg, 2005; Sawyer, 2012). Walau bagaimanapun, ia mungkin untuk mencari beberapa persamaan dari pelbagai definisi kreativiti. Semasa belajar, kanak-kanak yang kreatif, akan memberi tumpuan melalui personaliti kreatif. Beberapa kajian telah menekankan kebolehan transformasi proses kreatif. Secara tradisinya, kreativiti ditakrifkan melalui hasil daripada proses ini, iaitu, produk yang menjadi subjek kajian (Craft, 2005). Produk kreatif boleh menjadi ciptaan baru serta seni, teori, kemahiran atau kebiasaan. Idea kanak-kanak yang kreatif boleh menjadi suatu ciptaan yang kreatif.

Perspektif kreativiti kanak-kanak ini menekankan kepentingan alam sekitar dalam kreativiti; kreativiti digambarkan sebagai proses dan bahagian tidak dapat dipisahkan daripada budaya sekitarnya. Craft (2005) merumuskan definisi kreativiti memberi tumpuan sama ada di lokasi, pengeluaran atau kesan kreativiti. Kreativiti boleh nyata subjektif, secara kolektif atau secara aktif. Dalam erti kata lain, kreativiti juga boleh hadir dalam aktiviti kumpulan atau ketika proses kanak-kanak menghasilkan sesuatu. Definisi dijelaskan lagi ini melalui penghasilan idea-idea baru kanak-kanak yang dilihat dan berbeza sebahagian daripada kreativiti (Craft, 2005; Kudryavtsev, 2011).

Kreativiti adalah keupayaan untuk membangunkan sesuatu novelti dan menyesuaikan diri dengan situasi baru. Penyelesaian yang luar biasa di samping keaslian dilihat sebagai bahagian-bahagian yang tidak dapat dielakkan dalam

keaktiviti. (Hackbert, 2010). Walau bagaimanapun, cabaran yang timbul daripada teknologi moden dan ekonomi yang inovatif telah mewujudkan keperluan untuk pemikiran kreatif 'setiap hari'. Perubahan ini juga dilihat dalam bidang penyelidikan untuk kreativiti di mana kreativiti juga dikaji secara menyeluruh pada masa kini termasuk pada peringkat kanak-kanak (Brinkman, 2010; Craft, 2012; Lin, 2011; Paulus & Dzindolet, 2008; Sawyer, 2012).

Kreativiti kanak-kanak dan pembelajaran kreatif perkara yang berlaku setiap hari dalam kelas. Kreativiti perlu memenuhi dua kriteria: keaslian dan menambah makna yang besar kepada kumpulan tertentu. Kreativiti kanak-kanak sering berbeza daripada kreativiti orang dewasa kerana subjektiviti itu. Pendidik perlu melihat kreativiti kanak-kanak secara semula jadi. Kanak-kanak adalah sentiasa terbuka kepada pengalaman baru dan mempunyai tabiat yang berminat dalam segala-galanya yang baru (Lin, 2011). Subjektiviti merupakan satu ciri intrinsik kreativiti kanak-kanak. Sesuatu yang baru dalam idea-idea kreatif kanak-kanak tidak ditentukan oleh masyarakat, tetapi dengan ilmu pengetahuan yang terdahulu diperolehi daripada pengalaman atau pengamatan mereka (Kudryavtsev, 2011).

Pemikiran kreatif dikaji sebagai satu proses kognitif (Mumford, Medeiros & Partlow, 2012). Kajian Bacanlı (2011) menekankan ciri-ciri kognitif dalam pemikiran kreatif. Walaupun pemikiran kreatif termasuk kedua-dua ciri-ciri pemikiran kognitif dan afektif, sebahagian besar terdiri daripada pelbagai proses kognitif. Pemikiran kreatif mengandungi proses kompleks dan ilmu yang telah dipelajari sebagai satu proses melalui strategi berfikir, asas pengetahuan, atau dengan menggabungkan kedua-dua aspek. Penyelesaian masalah adalah contoh proses pemikiran kreatif (Mumford et al., 2012).

**Perkembangan Sosio-emosi.** Sosio-emosi adalah domain perkembangan yang penting kepada kecekapan menyediakan perkara asas bagi penguasaan pelbagai kemahiran yang penting kepada pencapaian akademik dan pencapaian (Denham, 2002; Jones & Bouffard, 2012). Hasilnya, kemahiran sosial-emosi adalah termasuk dalam satu set petunjuk kesediaan sekolah biasa yang digunakan untuk memacu peningkatan dalam keupayaan anak-anak taska untuk berjaya di tadika dan awal sekolah rendah.

Kecekapan sosio-emosi sering dianggap sebagai kategori terdiri daripada satu set kemahiran yang mempunyai ciri-ciri secara khusus. Salah satu ciri kemahiran ini adalah kecekapan sosial yang biasanya dicirikan bagi seorang kanak-kanak berinteraksi sosial dengan rakan sebaya dan orang dewasa secara berkesan (Fabes, Gaertner & Popp, 2006). Antara ciri sosio-emosi yang dilihat dalam aktiviti ialah, kanak-kanak mesti membina hubungan positif dengan orang lain, menyelaraskan dan berkomunikasi mengenai perasaan beliau dengan rakan-rakan sosial, dan mengenal pasti dan mengawal emosi dan tindakan beliau dalam persekitaran sosial.

Justeru, adalah penting untuk memahami kecekapan kanak-kanak yang berkembang dari masa ke masa dari segi sosioemosi, dan kognitif. Maka guru perlu memahami sosioemosi kanak-kanak secara asas seperti memberi perhatian, peringatan tentang aktiviti, memahami emosi dan menekankan peraturan, dan pengesanan isyarat tertentu kanak-kanak (Jones & Bouffard, 2012). Oleh itu, apabila mempertimbangkan pendekatan untuk mentaksir perkembangan sosioemosi kanak-kanak, ia adalah penting untuk mengambil kira beberapa ciri pada kanak-kanak secara keseluruhan serta menggunakan pendekatan yang sesuai semasa merangsang emosi mereka.

**Perkembangan Rohani.** Perkembangan rohani mempunyai kaitan dengan peringkat perkembangan kognitif (Elkind, 1970; Goldman, 1964; Boyatzis, 2005). Pada awal tahun 2000, banyak penyelidikan dalam perkembangan kognitif menggunakan

pendekatan untuk pembangunan agama secara khusus memberi tumpuan kepada hubungan antara membangunkan kognitif, sosial iaitu teori fikiran dan memahami konsep agama. Peralihan dalam tumpuan sebahagian besarnya didorong oleh perkaitan tentang sifat asas kognitif dengan perkembangan rohani. (Barrett, 2004; Boyer & Walker, 2000). Banyak penyelidikan semasa ke dalam pembangunan konsep agama telah beralih kepada selain daripada kognitif, seperti konsep agama sebagai penyesuaian dalaman dan peribadi diri (Atran, 2002; Bering, 2006; Boyer, 2001; Kirkpatrick, 2004)

Kajian Barrett & Richert (2003) menjelaskan kecenderungan kognitif tertentu yang mana apabila manusia dilahirkan, minda manusia "bersedia" untuk menerima konsep agama. Kajian konsep ketuhanan untuk perkembangan konsep rohani memberi tumpuan kepada pengesanan terhadap membangunkan pemahaman dalam fikiran kanak-kanak. Kajian Flavell (2004) pula menghuraikan dalam teori kerangka minda, bahawa perspektif seseorang itu adalah berbeza daripada yang lain, dan bahawa orang mempunyai pengetahuan yang berbeza tentang dunia. Dalam menimbang konsep agama secara khusus, terdapat kajian melihat konsep agama kanak-kanak iaitu daripada aspek pengetahuan, perspektif, dan kepercayaan Tuhan apabila kanak-kanak mula memahami batasan minda manusia. Kajian tersebut juga mendapati bahawa peringkat umur tertentu kanak-kanak mula mengaitkan kepercayaan, perspektif, pengetahuan kepada manusia dengan Tuhan (Barrett, Richert, & Dreisenga, 2001; Barrett, Newman, & Richert, 2003; Richert & Barrett, 2005).

Kajian Makris & Pnevmatikos (2007), mendapati bahawa kanak-kanak berusia 3 dan 4 tahun menunjukkan batasan tertentu terhadap kewujudan Tuhan, berbanding manusia. Di samping itu, Lane (2010) mendapati bahawa kanak-kanak berusia 2 tahun mempunyai kepercayaan kepada superhero. Gimenez Dasf, Guerrero & Harris (2005)

mendapati bahawa kanak-kanak berumur 5 tahun boleh memahami kematian dan juga boleh meniru sifat-sifat psikologi, fizikal untuk watak fantasi (Shtulman, 2008).

**Teori Perkembangan Piaget (1980).** Terdapat beberapa huraian isi Teori Piaget yang penting iaitu:

*Kanak-kanak adalah pelajar yang aktif dan bermotivasi.* Piaget telah memfokuskan penglibatan aktif adalah konsep asas dalam Teori Kognitif Piaget iaitu melalui pengalaman sebenar dengan persekitaran fizikal, kanak-kanak mengembangkan kecerdasannya. Manakala pembelajaran aktif adalah perkara asas dalam konstruktivisme. Pembelajaran aktif bermaksud kanak-kanak membina pengetahuan melalui aktiviti fizikal dan mental dan mereka sangat aktif terlibat dalam sesuatu aktiviti penyelesaian masalah.

Apa yang penting adalah rangsangan daripada guru untuk melihat tindakan yang akan diambil. Contohnya, blok kecil diberi kepada bayi berumur enam bulan. Dia akan cuba masukkan dalam mulut dan cuba memakannya. Jika blok diberi kepada kanak-kanak berumur tiga tahun, dia akan cuba menyusunnya. Kedua-dua kanak-kanak ini cuba melibatkan diri dengan benda fizikal dan menjadi aktif belajar. Penglibatan aktif ini adalah semulajadi untuk semua kanak-kanak dan ia adalah bahagian asas bagaimana mereka belajar. Melalui kajian ini, teori ini diaplikasikan dengan merangsang mereka melalui aktiviti dan peralatan yang digunakan semasa proses pembelajaran dan pemerhatian.

*Kanak-kanak lebih membina konstruk berbanding menyerap maklumat.* Piaget menyatakan, kanak-kanak membina kepercayaan dan kefahaman daripada pengalaman mereka. Dalam istilah lain, perkara yang kanak-kanak lakukan dan ketahui disusun sebagai skema iaitu mengumpulkan tindakan atau pemikiran yang sama yang berulang kali sebagai tindak balas kepada persekitaran. Contohnya, bagi kanak-kanak, menghisap puting adalah dimulut, maka skema tersebut digunakan

terhadap pelbagai objek lain dan membina konstruk baru dengan memasukkan jari atau biskut ke dalam mulut. Walau bagaimanapun, dengan meningkat umur, pengalaman kanak-kanak turut meningkat, maka skema akan berubah dan menjadi integrasi antara satu sama lain. Contohnya, kanak-kanak akan mula mengaitkan sesuatu dengan sesuatu yang lain. Kanak-kanak mengetahui kucing dan ayam adalah dalam kategori haiwan. Perkembangan konstruk maklumat dan kematangan berfikir membenarkan kanak-kanak berfikir secara kompleks dan lebih logik.

***Hubungan interaksi fizikal dan persekitaran sosial adalah asas kepada perkembangan Kognitif.*** Menurut Piaget, aktif bereksperimen dengan dunia fizikal adalah penting untuk perkembangan kognitif. Melalui penerokaan dan bermain dengan objek fizikal seperti bermain dengan pasir dan air, bermain permainan dengan bola dan bat menunjukkan kanak-kanak belajar alam semulajadi melalui ciri-ciri seperti jumlah dan berat atau mengaitkan prinsip yang berkaitan dengan tekanan dan graviti. Hubungan interaksi sosial adalah sama penting dengan perkembangan kognitif. Melalui interaksi dengan orang iaitu perbualan dan ketidaksefahaman seperti konflik dapat menyedarkan kanak-kanak bahawa individu yang berbeza mempunyai pandangan yang berlainan. Contohnya, boleh dilihat semasa bermain, setiap kanak-kanak mengambil alat permainan yang baginya adalah lebih menyeronokkan dan berbeza daripada kawan yang lain.

***Pendekatan Piaget sebagai alat pentaksiran.*** Pendekatan yang digunakan oleh Piaget adalah pendekatan klinikal di mana, orang yang lebih dewasa akan mengemukakan masalah atau tugas dan bertanyakan kepada kanak-kanak pelbagai soalan agar kanak-kanak memberi tindak balas terhadap masalah atau tugas tersebut. Pendekatan ini mempamerkan kebolehan kanak-kanak untuk berfikir dan menjelaskan semasa menyelesaikan masalah atau tugas dan secara tak langsung meningkatkan kemahiran berfikir secara logik (diSessa, 2007). Pendekatan ini selari dengan konsep

pentaksiran prestasi yang memerlukan kanak-kanak menyelesaikan tugas dan menunjukkan pengetahuan yang dimiliki melalui prestasi yang dipersembahkan.

*Pengalaman secara hands-on.* Piaget menekankan penerokaan terhadap persekitaran fizikal perlu berfokus dan menumpukan kepada pengalaman kanak-kanak. Mereka belajar melalui interaksi dengan alam semulajadi. Kajian mendapati pengalaman *hands-on* yang dialami kanak-kanak dengan objek maujud menjadi lebih berkesan apabila digabungkan dengan proses pengajaran agar dapat membantu kanak-kanak membuat rumusan terhadap pemerhatian mereka (Mayer, 2004). Ini menunjukkan kehadiran guru adalah sebagai pembimbing dan penghubung kanak-kanak dengan proses memahami sesuatu maklumat.

Berdasarkan Teori Piaget, peringkat praoperasi (2 hingga 7 tahun) adalah permulaan pemikiran simbolik. Pada waktu ini, perkembangan bahasa kanak-kanak meningkat dengan pantas kerana berfikir setiap perkara ada tujuan dan sebab dan pertambahan perbendaharaan kata atau pemahaman simbol membolehkan otak mereka berfikir tentang kepelbagaian objek dan peristiwa. Peringkat ini sesuai dengan umur kanak-kanak di taska yang mula meneroka persekitaran pembelajarannya. Selari dengan kajian ini, proses penilaian menggunakan item prestasi melibatkan pelbagai jenis bahan manipulatif, bahan konkrit untuk pembelajaran dan aktiviti serta domain bahasa dan komunikasi dan domain yang lain menyediakan tugas yang memberi peluang kanak-kanak melibatkan diri dalam pertuturan, mendengar, membaca dan menulis.

Kesimpulannya, kanak-kanak pada peringkat taska yang berumur 2 hingga 4 tahun adalah berada pada peringkat praoperasi yang merupakan peringkat kanak-kanak sudah boleh memahami objek dipersekitaran. Perkembangan kanak-kanak bukan sahaja meliputi satu domain, tetapi melibatkan semua domain perkembangan dan paling utama melalui interaksi antara domain bermula dengan tahap pengetahuan

dan kemahiran berfikir kanak-kanak yang berbeza antara satu sama lain. Semasa proses pentaksiran dijalankan, pelbagai masalah dan tugas diberikan mengikut domain untuk melihat tahap keupayaan kanak-kanak menyelesaikannya berdasarkan prestasi yang ditunjukkan. Penyelesaian masalah dan tugas adalah berkait rapat dan dikawal daya pemikiran seperti yang ditekankan dalam teori Piaget ini.

**Teori Perkembangan Kanak-kanak Vygotsky (1987).** Terdapat beberapa huraian isi Teori Vygotsky yang penting iaitu:

**Teori Sosio-budaya.** Jika Piaget menekankan kawalan daya pemikiran terhadap semua tindakan, Vygotsky pula lebih menekankan peranan orang dewasa yang memberi kesan kepada perkembangan kognitif kanak-kanak secara sistematik. Secara umumnya, pengajaran dan bimbingan orang dewasa adalah penting kepada peningkatan perkembangan kognitif kerana faktor sosial dan budaya memberi kesan kepada perkembangan kognitif kanak-kanak. Perspektif ini dikenali sebagai teori Sosio-budaya Vygotsky. Vygotsky mengkaji jenis tugas yang hanya boleh diselesaikan dengan bimbingan orang dewasa. Selain itu, kanak-kanak akan diberi tugas yang lebih sukar dan melihat maklum balas daripada kanak-kanak yang memerlukan bantuan atau kemungkinan tugas tersebut boleh dilakukannya sendiri. Walau bagaimanapun, menurut Vygotsky, kadar tugas yang boleh diselesaikan oleh dua orang kanak-kanak yang berlainan adalah berbeza dengan salah seorang kanak-kanak tersebut mempunyai kebolehan untuk melakukan tugas tersebut pada tahap umur 12 tahun dan seorang lagi berjaya melakukannya hanya pada ketika usianya 9 tahun.

**Alat fizikal dan kognitif menjadikan kehidupan seharian kanak-kanak lebih produktif dan efisien.** Selain peranan orang dewasa, peralatan tertentu juga boleh membantu kanak-kanak berhadapan dengan pelbagai tugas dan masalah. Contohnya, gunting, komputer dan objek fizikal. Mengikut pandangan Vygotsky, alat



kognitif meningkatkan kebolehan berfikir kanak-kanak. Alat kognitif yang dimaksudkan adalah bahasa. Untuk orang dewasa, pemikiran dan bahasa adalah berhubung antara satu sama lain. Contohnya, apabila berfikir mengenai kelengkapan dapur, ia mengandungi sudu, peti sejuk dan akan dituturkan menggunakan medium bahasa. Walau bagaimanapun, pemikiran dan bahasa kanak-kanak mempunyai fungsi yang berbeza dan tersendiri. Pada peringkat awal kanak-kanak, pemikiran terjadi secara berasing dengan bahasa, dan apabila bahasa digunakan ia adalah asas yang memberi kepada makna sesebuah komunikasi.

Bermula umur dua tahun, kanak-kanak mula meluahkan pemikiran mereka semasa mereka bercakap dan mereka mula berfikir mengenai perkataan. Kadang-kadang mereka akan bercakap sendiri (*self-talk*). Vygotsky menerangkan bahasa bercakap sendiri adalah fungsi yang penting dalam perkembangan kognitif. Bercakap dengan diri sendiri membantu kanak-kanak belajar untuk mengawal diri dan menunjukkan tingkah laku semasa melakukan tugas yang sukar.

***Kanak-kanak boleh melakukan tugas yang lebih sukar apabila dibantu oleh individu yang lebih cekap.*** Vygotsky membezakan dua jenis kebolehan kanak-kanak yang memberikan ciri-ciri kemahiran tertentu dalam peringkat perkembangan. Pertama, tahap perkembangan sebenar kanak-kanak iaitu pada tahap kanak-kanak dapat melakukan dengan sendiri, tanpa bantuan daripada sesiapa. Kedua, tahap potensi perkembangan kanak-kanak iaitu pada tahap kanak-kanak dapat melakukannya dengan bantuan individu yang lebih cekap. Penilaian perlu dilakukan terhadap kebolehan pada kedua-dua tahap iaitu semasa bersendirian atau dengan bantuan. Ia bertujuan untuk mendapatkan keadaan perkembangan kognitif kanak-kanak yang sebenar. Kajian ini akan dilaksanakan dengan meminta kanak-kanak melakukan tugas dan melihat tahap kebolehan samada perlu diberi bantuan atau sebaliknya sebelum memberikan skor.

*Tugas yang mencabar mencapai perkembangan kognitif ke tahap maksimum.* Mengikut Teori Vygotsky, tugas yang mencabar menunjukkan perkembangan kognitif yang maksimum. Terdapat beberapa peringkat tugas di mana kanak-kanak dapat melakukannya sendiri dengan bantuan, namun menunjukkan hasil yang berbeza selepas bimbingan diberikan. Peningkatan perbezaan perkembangan ini merujuk kepada *Zone of Proximal Development (ZPD)*. Zon ini penting untuk menentukan potensi perkembangan dan kebolehan sebenar kanak-kanak melakukan sesuatu aktiviti dan mengetahui tahap kesukaran aktiviti tersebut.

Pada peringkat ini, pencapaian kanak-kanak boleh dicatatkan kerana setiap tugas yang diberikan akan menentukan kebolehan sebenar kanak-kanak sebelum dan selepas diberi bimbingan bergantung kepada kesukaran tugas itu atau sememangnya boleh melakukan dengan cekap tanpa sebarang bimbingan yang menunjukkan tugas itu mudah untuknya. Bimbingan dalam ZPD dikenali sebagai *Scaffolding* dan ia adalah komponen utama dalam pengajaran guru kerana guru memainkan peranan penting dalam membimbing dan berinteraksi dalam pembelajaran kanak-kanak. Kajian ini, akan melihat kepelbagaian peringkat perkembangan kanak-kanak yang selari dengan ZPD semasa pembelajaran dalam semua domain perkembangan untuk menentukan tahap kebolehan kanak-kanak, tahap kesukaran tugas dan peranan guru dalam melaksanakan aktiviti pengajaran agar kepelbagaian peringkat perkembangan ini dapat diperhatikan dan dinilai oleh penilai.

***Konsep Intersubjectivity.*** Aspek penting yang ditekankan dalam teori ini adalah konsep *intersubjectivity* berdasarkan kepelbagaian perkembangan kanak-kanak. Kanak-kanak adalah individu. Maka hasil penilaian melihat perkembangan kanak-kanak tidak dibandingkan antara satu sama lain kerana konsep *intersubjectivity*. Setiap kanak-kanak yang hendak melakukan sesuatu tugas, menyelesaikan masalah atau berkomunikasi membawa bersama persepsi subjektif yang tersendiri. Semasa proses

interaksi, setiap tindakan terhasil akan berbeza dengan kanak-kanak yang lain. Maka, peranan guru mengorganisasi persekitaran pembelajaran dan menyediakan peluang berhubungan dengan cara yang sesuai di dalam kelas dapat menyokong *intersubjectivity* setiap kanak-kanak seterusnya penilaian dilakukan dengan lebih berkesan kerana setiap kanak-kanak adalah berbeza perkembangannya.

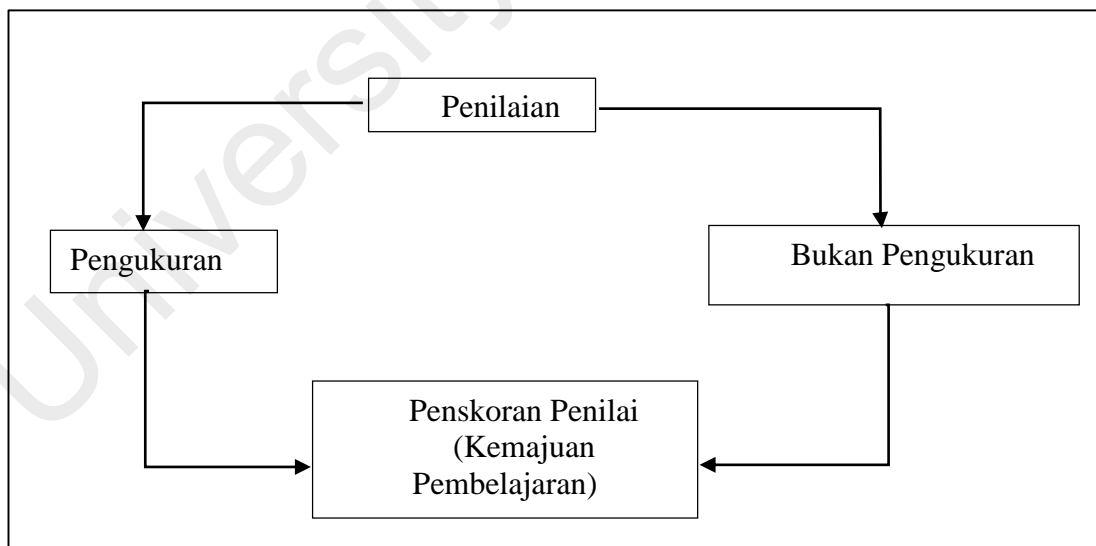
Kesimpulannya, pendekatan teori yang digunakan sebagai garis panduan semasa proses pengujian dilaksanakan. Teori Piaget menekankan aspek daya pemikiran semasa melaksanakan tugas pembelajaran yang dilihat berkaitan semasa proses mentaksir kanak-kanak. Manakala Vygotsky memberi garis panduan terhadap bentuk perkembangan yang akan dicapai kanak-kanak berdasarkan usaha yang ditunjukkan. Justeru, gabungan kedua-dua teori ini membantu pentaksiran secara holistik dapat dilaksanakan sekaligus menunjukkan tahap kebolehan kanak-kanak berada pada tahap yang berbeza mengikut kepelbagaian perbezaan perkembangan yang dimiliki setiap kanak-kanak.

**Rumusan teori perkembangan kanak-kanak.** Fokus seseorang kanak-kanak mengikut teori Piaget dan Vygotsky adalah untuk menyelesaikan masalah dalam situasi tertentu dengan menggunakan domain-domain pembelajaran yang ada pada mereka. Penyelidik berhasrat menguji instrumen yang mengandungi item-item mentaksir prestasi dan kemahiran seseorang kanak-kanak untuk mengesan tahap kebolehannya, tahap kesukaran item dan domain perkembangan serta pengaruh individu atau penilai terhadap kanak-kanak sebagaimana dinyatakan dalam kedua-dua teori tadi supaya pengesanan mencakupi sebahagian besar pencapaian kanak-kanak berdasarkan keenam-enam domain.

**Konsep penilaian dan pengukuran.** Penilaian adalah istilah umum yang merangkumi rangkaian penuh prosedur yang digunakan untuk mendapatkan maklumat mengenai pembelajaran pelajar (pemerhatian, penilaian prestasi atau projek, ujian

kertas dan pensil) dan melibatkan penskoran penilai. Pengukuran adalah merupakan nilai nombor kepada keputusan ujian yang mengikut format tertentu (contoh, mengira jawapan yang betul untuk memberi mata untuk aspek khusus kepada esei). Penilaian adalah satu istilah yang lebih komprehensif dan inklusif daripada pengukuran. Pengukuran adalah terhad kepada huraian kuantitatif pelajar, iaitu, hasil daripada pengukuran sentiasa dinyatakan dalam nombor (contoh, Alex mendapat betul 35 daripada 40 penyelesaian masalah matematik). Penilaian, di sudut lain, boleh meliputi kedua-dua penerangan kuantitatif (ukuran) dan penerangan kualitatif (bukan pengukuran) pelajar. Selain itu, penilaian yang melibatkan penskoran oleh penilai perlunya keputusan.

Rajah 2.1 menunjukkan aspek penilaian yang komprehensif dan peranan aspek pengukuran dan bukan pengukuran dalam proses penilaian. Seperti yang dinyatakan dalam rajah, penilaian boleh atau tidak berasaskan kepada pengukuran. Jika ia melibatkan pengukuran, ia akan melibatkan huraian secara kuantitatif.



Rajah 2.1. Proses Penilaian oleh Miller, Linn & Gronlund (2013)

**Konsep pentaksiran kanak-kanak.** Pentaksiran kanak-kanak adalah proses pengumpulan maklumat tentang kemajuan dan perkembangan mereka. Proses pentaksiran dilakukan melalui pemerhatian dengan menggunakan pelbagai kaedah.

Dengan adanya pentaksiran, guru dapat mengenal pasti potensi, kekuatan atau kelemahan kanak-kanak dalam pembelajaran dari semasa ke semasa. Seterusnya, guru dapat melakukan tindakan susulan jika perlu. Guru juga dapat mengembangkan potensi kanak-kanak ke tahap maksimum dalam domain penting, iaitu kognitif, afektif dan psikomotor.

Golongan awal kanak-kanak ialah mereka yang sukar dinilai secara jitu kerana aktiviti yang dilakukan mereka, tumpuan yang singkat dan pencapaian yang tidak konsisten dalam persekitaran baharu. Faktor lain yang juga boleh mempengaruhi pencapaian kanak-kanak termasuklah perbezaan budaya, bahasa, baka serta interaksi ibu bapa dengan anak-anak mereka. Untuk itu, kanak-kanak amat memerlukan kaedah pentaksiran yang berkesan dalam mengenal pasti kemajuan mereka dengan menggunakan cara yang sah dan boleh dipercayai (Mariani, 2011).

Pentaksiran amat berkesan dilakukan semasa keadaan yang tidak formal atau bersahaja. Selain itu, penilaian terhadap kanak-kanak dalam konteks komuniti dan sistem interaksi sosial mereka dapat menyumbang kepada keberkesanan dalam penilaian yang dilakukan itu. Menilai perkembangan kanak-kanak melibatkan pemerhatian, rakaman serta dokumentasi kegiatan dan permainan yang dijalankan oleh kanak-kanak. Hal ini merupakan proses yang berterusan dan bertumpu pada perkara yang dilakukan oleh kanak-kanak dan cara perkara itu dilakukan (Meisels, 1995). Menurut Seefeldt (1990), kanak-kanak mempunyai tingkah laku yang terhad untuk ditafsirkan dan tidak dapat menggunakan bahasa dengan tepat bagi menerangkan pengalaman yang dialaminya. Oleh itu, pemerhatian ialah kaedah terbaik untuk menilai perkembangan mereka kerana kanak-kanak meluahkan perasaan, idea, pengetahuan dan emosi mereka melalui tingkah laku.

**Teori Pengukuran.** Teori pengukuran sangat penting dalam menyelesaikan permasalahan dalam pembinaan instrumen. Teori pengukuran yang terlibat adalah Teori Ujian Klasikal (TUK) dan Teori Respons Item (TRI). Kedua - dua teori ini saling melengkapi dan mempunyai beberapa andaian yang berbeza. Penyelidikan dalam pengukuran telah mengubah penekanan ujian ke arah teori yang berfokuskan ujian kepada yang berfokuskan item. Pendekatan ini sebut sebagai Teori Trait Pendam atau Teori Keluk Ciri Item yang kini dipanggil Teori Respons Item (TRI) (Hambleton & Swaminathan, 1985; Hulin, Drasgow, & Parson, 1983). TRI sering digunakan dalam bidang pengujian (Siti Rahayah, 2008).

**Teori Ujian Klasik.** Teori Ujian Klasikal (TUK) digunakan dalam pembinaan ujian yang berdasarkan kepada indeks kesukaran dan indeks diskriminasi item. TUK menghubungkan tiga konsep utama iaitu skor ujian, skor benar dan skor ralat (Miller, 2013). Teori dikenali sebagai model Skor Benar dan diwakili persamaan berikut:

$$X = T + E$$

X = skor pemerhatian

T = skor benar

E = ralat rawak

Skor benar ini menggambarkan nilai kebolehan atau sikap individu yang sebenar. Teori ini mengandaikan tret adalah malar dan variasi dalam skor pemerhatian adalah disebabkan oleh ralat rawak yang dipengaruhi oleh faktor seperti meneka jawapan dan faktor kelesuan (Crocker & Algina, 1986). Apabila nilai ralat adalah sifar, maka skor pemerhatian adalah sama dengan skor benar.

TUK mempunyai beberapa kelemahan iaitu (i) TUK mengandaikan ralat pengukuran disebarkan secara normal dan sama untuk semua tahap skor. TUK menganggap semua item mempunyai nilai yang sama dan menterjemahkannya dal am

bentuk purata, peratusan dan kebarangkalian (Bond & Fox, 2007; Wright & Stone, 1979); (ii) skala kebolehpercayaan item bergantung kepada jumlah sampel dan skor bergantung kepada bilangan item dalam instrumen, keputusan akan berubah jika sampel berubah; (iii) tidak dapat menganggar tahap kebolehan pengambil ujian kerana TUK hanya melibatkan item betul dan salah saja tanpa mengambilkira keupayaan seseorang individu ke atas sesuatu item dan (iv) format yang berbeza - beza dalam TUK menyebabkan jumlah skor ujian tidak seimbang kerana kaedah yang terhad dan statistiknya bergantung kepada sampel.

Pendekatan Likert sepertinya kelihatan mudah bagi menyelesaikan permasalahan yang berkaitan dengan pengukuran yang dihadapi, tetapi sebenarnya berlaku pelanggaran terhadap beberapa hukum asas matematik secara terang - terangan (Azrilah Abdul Aziz, Mohd Saifudin Masodi & Azami Zaharim, 2013). Berdasarkan andaian dan kemampuan pengukuran TUK pada skala rating, salah satu limitasi yang dikenal pasti adalah jenis data yang digunakan. Analisis statistik inferensi seperti korelasi, regresi, ANOVA dan sebagainya memerlukan data jenis selang dan nisbah sahaja yang boleh digunakan untuk pengukuran sebenar. Namun, syarat ini tidak dipatuhi kerana pengukuran skala Likert adalah berbentuk data ordinal. Skala Likert adalah merupakan data jenis ordinal dan bersifat pangkatan. Pangkatan meningkat seiring dengan tret. Anggapan bahawa pertambahan pangkatan menunjukkan peningkatan yang sama antara suatu skala dengan skala yang lain adalah tidak tepat (Azrilah et al, 2013). Perbezaan skor mentah antara pasangan mata tidak semestinya membayangkan jumlah skor yang sama bagi konstruk yang dikaji (Smith, Conrad, Chang, & Piazza, 2002).

Contohnya bagi skala Likert lima mata, perbezaan antara skala setuju dan sangat setuju mungkin tidak sama dengan perbezaan antara tidak setuju dengan sangat tidak setuju atau tidak setuju dengan neutral. Likert sering dianggap bersifat bersusur

galur (kontinum) walaupun andaian tersebut didapati meleset sama sekali (Azrilah et al., 2013). Skor mentah skala Likert bagi analisis TUK seringkali disalah anggap sebagai data selang dan telah disalah gunakan dalam prosedur statistikal berparametrik (Bode & Wright, 1999). Justeru, pengkaji bakal membuat kesimpulan yang salah tentang signifikan kajian yang dibuat jika berlaku kesilapan memilih teknik statistik. Model Rasch boleh mengatasinya dengan mentransformasikan data ordinal (skala likert) ini kepada data interval melalui nilai ukuran kesukaran (Bond & Fox, 2007; Wright & Stone, 1979).

**Teori Respons Item.** Salah satu teori alternatif untuk memperbaiki kelemahan yang terdapat pada TUK adalah Teori Respon Item (TRI). TRI merupakan satu pendekatan yang digunakan untuk mendapatkan gambaran tentang fungsi item yang dapat menjangkakan prestasi individu ke atas sesuatu item ujian (Crocker & Algina, 1986). TRI membentuk satu keluarga model matematik yang menghuraikan perlakuan item dalam sesuatu ujian berdasarkan Fungsi Respons Item (FRI). FRI adalah fungsi matematik yang menghubungkan kebarangkalian respons yang betul bagi sesuatu item dan kebolehan item mengukur trait terpendam. Konsep ini juga disebut sebagai ICC atau Keluk Ciri Item. Terdapat empat model TRI seperti Model 1, 2, 3 dan 4 parameter yang boleh menghuraikan FRI. Model Rasch atau satu parameter logistik (1 - PLM) digunakan dalam kebanyakan pengujian psikologi, model dua parameter logistik (2 - PLM) bagi ujian soalan ringkas dan model tiga parameter logistik (3 - PLM) bagi soalan objektif pelbagai pilihan. Model empat parameter jarang digunakan dalam mana - mana aplikasi (Abu Bakar & Bhasah, 2008; Siti Rahayah, 2008). Model Rasch atau satu parameter logistik (1 - PLM) biasanya digunakan dalam pengujian instrumen psikologikal.



TRI mempunyai semua ciri - ciri yang difikirkan mampu mengatasi semua masalah pada TUK adalah seperti (i) ciri-ciri item tidak lagi bergantung pada sampel (kumpulan yang cerdik atau kurang cerdik); (ii) skor yang menghuraikan kebolehan pelajar dan tidak lagi bergantung pada ujian (ujian sukar atau mudah); (iii) model dapat membekalkan maklumat sehingga paras item (bukan hanya pada paras ujian); (iv) model tidak memerlukan ujian selari untuk mentaksirkan kebolehpercayaannya (Siti Rahayah, 2008). Model Rasch atau model trait laten menghasilkan kerangka alternatif untuk memberi pemahaman pada pengukuran serta strategi alternatif bagi menilai kualiti instrumen (Kimberlin & Winterstein, 2008).

**Perbandingan di antara TUK dan TRI.** TUK dan TRI mempunyai objektif yang sama untuk menyelesaikan permasalahan dalam membina instrumen. Kedua - dua teori ini saling melengkapi antara satu sama lain dan mempunyai kelemahan dan kekuatan tersendiri. Jadual 2.3 menunjukkan perbandingan utama antara TUK dan TRI (Siti Rahayah, 2008).

Jadual 2.3  
*Perbandingan Utama di antara TUK dan TRI*

Teori	TUK	TRI
Model	Linear $X = T + E$	Tak linear
Tahap	Ujian	Item
Ralat Pengukuran	Ralat (E) = $X - T$	Ralat = Respon pemerhatian - Respon diramalkan (Ralat = respon diperolehi - respon ramalan)
Interprestasi skor	$X + SEM$	Rasch : logits $\pm$ residual TRI: $0 \pm$ ralat Skor menunjukkan kebarangkalian memberi maklum balas (menjawab) dengan betul bagi sesuatu item model terpendam
Hubungan	Tidak dinyatakan	Keluk ciri item (ICC)
Keupayaan	Skor ujian (atau anggaran skor sebenar dilaporkan pada skala skor ujian)	Keupayaan skor dilaporkan pada skala - $\infty$ ke $+\infty$
Invarian bagi item dan individu	Tiada - parameter item dan individu adalah saling bergantung	Ya — parameter item dan individu adalah bebas atau tidak bergantung pada sampel sekiranya model sesuai untuk data kajian
Saiz sampel	Secara umum di antara 200 ke 500	Bergantung pada model TRI yang digunakan tetapi umumnya tidak memerlukan sampel yang besar

Kekuatan	Lemah (mudah untuk memenuhi keperluan data ujian) $E(e) = 0$ $P_{te} = 0$ $P_{ele2} = 0$	Mempunyai kekuatan (lebih sukar untuk memenuhi keperluan data ujian). Unidimensi (bergantung antara item atau bilangan tret pendam yang diperlukan untuk mencapai kebebasan setempat iaitu bebas antara item — item pada tahap keupayaan)
----------	---	--

### Model Rasch

**Andaian Rasch.** Model Rasch diperkenalkan oleh Georg Rasch iaitu seorang pakar Matematik dari Denmark pada tahun 1960. Model ini juga disebut sebagai model satu parameter iaitu parameter kesukaran item dan tahap kebolehan *person* sahaja. Model Rasch didasari oleh Teori Respon Item (TRI). TRI juga disebut sebagai Teori Trait Pendam atau Teori Keluk Ciri Item (ICC). Asas bagi model ini adalah untuk mengasingkan kebolehan individu dari kualiti instrumen. Model ini menganggap bahawa respons individu terhadap sesuatu item hanya dipengaruhi oleh kebolehan individu dan kesukaran item sahaja (Bond & Fox, 2007).

Model Rasch mendapat perhatian ramai pengkaji dalam dan luar negara yang menggunakan pendekatan ini dalam kajian mereka terutamanya bagi pengesahan kesahan dan kebolehpercayaan instrumen yang baru dibina (Arasinah, Bakar, Ramlah, Soaib, & Zaliza, 2014; Dur et al., 2014; Franchignoni, Giordano, Ronconi, Rabini, & Ferriero, 2014; Hecimovich, Styles, & Volet, 2014; Mashitoh, 2013; McCreary et al., 2013; Mofreh, Ghafar, Omar, Mosaku, & Ma'rif, 2014; Norasmah, Suria, Haliza, & Haryaty, 2014; Rohaya & Mohd Najib, 2008; Siti Rahayah, Farhana, Ayesha, & Nur Aidah, 2011; Sondergeld & Johnson, 2014; Syed Muhd Kamal, Ahmad, & Syed Muhamad Dawilah, 2014; Mohd Effendi, 2015). Model Rasch sering menjadi pilihan kerana mempunyai beberapa kelebihan antaranya faktor kelinearan, bebas, objektif, menyeluruh dan inferens boleh ditunjukkan dengan jelas (Wright & Stone, 1979).

Model Rasch mengandaikan dalam setiap item dibina wujud parameter kesukaran sahaja. Model Rasch menjelaskan bahawa seseorang individu mempunyai

peluang 50:50 untuk berjaya pada soalan yang diberi pada kesukaran 0.00 logits. Apabila item bertambah sukar, kebarangkalian untuk berjaya akan berkurang atau peluang kejayaan semakin tipis (Rasch, 1961, 1980).

Turutan penilaian dalam model Rasch adalah berbeza mengikut pengkaji lepas. Terdapat pengkaji yang menyemak skala terlebih dahulu (Cox et al., 2014; Franchignoni et al., 2014; Hill & Koekemoer, 2013; Lerdal, Fagermoen, Bonsaksen, Gay, & Kottorp, 2014). Ada juga pengkaji yang memulakannya dengan keserasian item (Apple, 2013; Lopes, Prieto, & Delgado, 2014; Ramakrisnan, Azizah, & Nor Faedah, 2013; Sondergeld & Johnson, 2014) mahupun unidimensionaliti (McCreary et al., 2013; Mofreh et al., 2014). Tiada keseragaman antara pengkaji lepas apabila melakukan penilaian model Rasch. Contohnya, terdapat pengkaji yang tidak mempertimbangkan sesetengah analisis Rasch seperti kebebasan setempat dan DIF (Mofreh et al., 2014; Prasanna Ramakrisnan et al., 2013b; Sondergeld & Johnson, 2014). Sebenarnya ia banyak bergantung kepada objektif kajian dan jenis analisis Rasch yang diperlukan dalam kajian.

Fisher (2006) menegaskan bahawa prosedur pertama yang perlu dilakukan adalah mengenal pasti item yang tidak memenuhi model Rasch (tidak fit) dengan nilai MNSQ atau statistik kesesuaian yang seragam (Zstd) bagi setiap item. Analisis seterusnya boleh dijalankan selepas menjalankan analisis awal tadi. Ini bermaksud analisis keserasian item dipilih sebagai prosedur pertama. Fisher (2006) kemudiannya menjelaskan bahawa analisis perlu dijalankan untuk mengetahui sama ada terdapat item yang bertanyakan dua atau lebih soalan pada satu masa. Pandangan ini selari dengan kenyataan bahawa salah satu pertimbangan paling penting dalam pembangunan sesuatu instrumen adalah bersifat unidimensi, iaitu menganggap item-item dalam instrumen adalah mengukur satu keupayaan tunggal (Wright & Masters, 1982). Pengkaji lepas seperti Kim dan Hong (2004) turut menyifatkan

unidimensionaliti sebagai andaian paling asas dan paling kritikal dalam model Rasch yang mesti dipenuhi terlebih dahulu.

Penilaian item adalah berdasarkan kepada prinsip bahawa terdapat beberapa andaian yang mesti dipenuhi sebelum TRI dapat digunakan dalam beberapa aplikasi. Andaian tersebut adalah statistik kesesuaian, unidimensionaliti dan kebebasan setempat (Siti Rahayah, 2008). Siti Rahayah (2008) menjelaskan pengkaji mesti menaksirkan kesesuaian model TRI kepada data yang hendak digunakan. Walaupun syarat ini melibatkan proses pengguguran item, namun ia boleh membantu meningkatkan unidimensionaliti item bagi konstruk yang ingin diukur (Balsamo et al., 2014). Justeru, beberapa syarat asas mesti dipatuhi dalam model Rasch seperti andaian utamanya iaitu (i) kefungsian skala (ii) keserasian item, (ii) unidimensionaliti dan (iii) kebebasan setempat selain kriteria tambahan seperti (iv) polariti item dan (v) kebezaan kefungsian item. Kajian juga meneliti (vi) pemetaan - item individu, (vii) semakan skala pemeringkatan serta (viii) indeks kebolehppercayaan dan indeks pengasingan.

***Kefungsian Skala.*** Sebanyak enam kriteria yang sesuai digunakan untuk menilai keberkesanan sesuatu skala pengukuran dengan model Rasch (Linacre, 2002). Pertamanya, setiap kategori mesti mempunyai sekurang - kurangnya 10 pemerhatian. Ini penting kerana jika nilai frekuensi kategori adalah rendah, jangkaan ketepatan skor Rasch boleh mendatangkan bermasalah. Kedua, adalah melalui bentuk lengkung kebarangkalian yang sepatutnya puncak pada setiap kategori. Skor pengiraan Rasch mesti menunjukkan taburan seragam bagi setiap kategori. Ketiganya, nilai ukuran purata bagi kategori mesti meningkat seiring dengan skala kategori. Contohnya, skala likert empat mesti mempunyai nilai ukuran yang lebih tinggi daripada kategori yang lebih rendah (likert tiga, dua dan satu).

Keempat, nilai Outfit MNSQ mesti kurang daripada dua. Statistik ukuran Outfit menjelaskan tentang varians. Skor melebihi dua menunjukkan tahap gangguan

dengan varians tidak dijelaskan. Kelima, tentukan sekatan (threshold) mesti meningkat seiring dengan kategori skala rating. Kriteria ini melibatkan kemampuan menjangka skor responden. Peningkatan ini menambahkan lagi kebarangkalian responden diperhatikan pada skala rating yang lebih tinggi. Semakan dilakukan pada jenis kategori skala dan pemerinkatan skala yang digunakan. Semakan pada jenis kategori skala yang digunakan merujuk kepada nilai ukuran purata yang meningkat secara sekata mengikut label kategori skala Likert. Peningkatan sekata menunjukkan pola respons yang normal dan seragam membuktikan kategori skala memuaskan. Keenam, kategori sekatan mesti melebihi satu tetapi tidak melebihi nilai lima bagi skala likert empat mata (Linacre, 2002). Sekatan Rasch Andrich digunakan bagi menunjukkan titik perubahan seseorang individu semasa proses pembuatan keputusan bergerak dari satu skala ke skala bersebelahan dan seterusnya. Sekiranya perbezaan kurang daripada satu, maka rating yang diandaikan perlu dirangkumkan dan jika melebihi lima, maka rating mestilah dipisahkan. Keenam - enam kriteria ini adakalanya tidak dipenuhi sepenuhnya, namun dapatan yang paling memuaskan akan dipertimbangkan untuk menilai keberkesanan skala rating.

***Keserasian atau kesepadanan item (fit).*** Statistik keserasian item menunjukkan sejauh mana sesuatu data adalah berpadanan dengan model Rasch (Siti Rahayah, 2008). Ia membantu pembina instrumen untuk memutuskan akan kesesuaian sesuatu item (Wright & Masters, 1982). Dalam model Rasch, dua jenis statistik keserasian yang diberikan adalah analisis purata kuasa dua (MNSQ) dan statistik kesesuaian yang seragam (Zstd) yang boleh digunakan bagi mengesan sama ada data menunjukkan percanggahan dengan model. Nilai jangkaan MNSQ yang sempurna adalah bernilai 1.00 logits.

Terdapat penjelasan yang tidak konklusif bagi penilaian Infit MNSQ atau Outfit MNSQ. Bond & Fox (2007) menegaskan bahawa nilai Outfit MNSQ tidak

diberikan pemberat kerana ia lebih mudah dipengaruhi oleh responden yang merasakan sesuatu item tersebut terlalu mudah atau terlalu sukar untuk dipersetujui. Ia juga tidak memberikan informasi berguna tentang majoriti responden. Pengkaji biasanya memberikan lebih perhatian kepada Infit MNSQ kerana lebih mengenal pasti kualiti item yang digunakan pada majoriti responden. Azrilah et al. (2013) pula menjelaskan bahawa indeks Outfit MNSQ perlu di berikan perhatian kerana ia lebih mudah diatasi berbanding dengan masalah Infit MNSQ yang sukar didiagnosis dan diatasi. Maka ia boleh diabaikan sama sekali.

Justeru, indeks kedua-dua Infit MNSQ dan Outfit MNSQ perlu diberi perhatian bagi menentukan kesepadanan item atau item fit yang mengukur sesuatu konstruk atau variabel laten. Pengkaji lepas menetapkan julat yang perlu dipatuhi bagi menguji andaian keserasian item contohnya nilai MNSQ item harus berada pada julat 0.77 logits - 1.30 logits (Fisher, 2007). Nilai Zstd pula menunjukkan kepentingan data. Ia adalah Statistik Kesesuaian Min Kuasa Dua yang menganggarkan teoritikal min dan taburan varian. Nilai Zstd antara - 2.0 hingga 2.0 adalah nilai diterima (Bond & Fox, 2007). Namun begitu, jika MNSQ diterima, nilai Zstd boleh diabaikan (Linacre, 2005).

***Unidimensionaliti.*** Unidimensionaliti bermaksud item-item dalam instrumen mengukur satu keupayaan tunggal (Wright & Masters, 1982). Terdapat empat kriteria yang digunakan bagi menguji andaian unidimensionaliti iaitu *Principal Component Analysis of Residual* (PCA) sekurang-kurangnya 20 peratus sebagai pertimbangan pengukuran dimensi Rasch yang paling minima (Conrad, Dennis, & Funk, 2012); 40 peratus *raw variance explained by measures* adalah tanda ukuran instrumen unidimensionaliti yang baik (Azrilah et al. 2013). Tahap gangguan item yang diukur atau varian yang tidak dijelaskan dalam kontras satu yang kurang 10 peratus (Eakman, 2012; Linacre, 2007), nilai kadar minimum nisbah 3:1 iaitu kadar nisbah varians yang dijelaskan oleh ukuran dengan varians komponen prinsipal pertama (Conrad et al.,

2012; Embretson & Reise, 2000). Nilai eigen kurang daripada lima menunjukkan tidak wujud dengan jelas dimensi kedua (Linacre, 2005). Item yang mempunyai kebebasan setempat bermaksud respon kebolehan individu kepada mana - mana item adalah tidak berkait dengan respon item lain dalam konstruk lain.

Nilai polariti item membuktikan item - item mengukur konstruk yang sama dan semua item mengukur sub konstruk tunggal (Bond & Fox, 2007). Analisis ini juga menunjukkan terdapatnya item atau individu yang memberi respon bercanggah dengan variabel (Linacre, 2005). Nilai polariti item (PTMEA) yang diperolehi sebaiknya memiliki nilai positif dan melebihi 0.3 logits (Bond & Fox, 2007; Nunnally & Bernstein, 1994; Wu & Adams, 2007). Julat nilai polariti yang baik bagi sesuatu instrumen adalah 0.3 logits hingga 0.6 logits (Bond & Fox, 2007).

**Nilai kebolehppercayaan.** Kebolehppercayaan dalam model Rasch menunjukkan kebolehppercayaan kebolehan individu yang bertujuan untuk menentukan konsistensi bagi respon seseorang individu. Kebolehppercayaan kesukaran item pula bertujuan untuk menganggar sebaik mana sesuatu item itu bersifat koheren atau berhubungan antara satu sama lain (Wright & Masters, 2002). Fisher (2007) menegaskan nilai kebolehppercayaan melebihi 0.94 sebagai cemerlang, 0.93 hingga 0.91 sebagai amat baik dan 0.90 hingga 0.81 dianggap baik. Indeks kebolehppercayaan individu melebihi 0.8 dengan kebolehppercayaan item melebihi 0.9 juga membuktikan sampel adalah mencukupi (Linacre, 2012). Kebolehppercayaan Cronbach Alfa turut diperhatikan dengan nilai melebihi 0.80 sebagai baik (Mohamad Najib, 2009; Othman, 2013; Sekaran & Bougie, 2011).

**Indeks pengasingan.** Indeks pengasingan dalam model Rasch dibahagikan kepada dua iaitu indeks pengasingan individu yang bertujuan untuk menganggarkan kebolehan sesuatu instrumen untuk membahagikan individu kepada beberapa tahap dalam konstruk yang diukur. Indeks pengasingan item pula bertujuan untuk

menganggarkan kebolehan individu dengan mengasingkan kesukaran item kepada beberapa tahap dalam konstruk yang diukur (Wright & Masters, 2002). Nilai indeks pengasingan individu dan item melebihi dua adalah baik (Jones & Fox, 1998). Indeks menunjukkan bilangan strata kebolehan individu dan kesukaran item. Strata melebihi dua dianggap membentuk variabel yang tersebar dengan baik dan kedudukan item pada skala logits mempunyai kebolehpercayaan tinggi (Siti Rahayah, 2008). Semakin besar nilai indeks pengasingan, maka kualiti instrumen pada responden dan kelompok item adalah semakin bagus (Bambang & Wahyu, 2014).

*Pemetaan Individu, Penilai dan Item (Huraian Rajah).* Pemetaan item - individu menunjukkan taburan kesukaran item - item instrumen yang dipadankan dengan taburan kebolehan pelajar di sepanjang skala logits pada satu kontinum pengukuran iaitu dari aras paling mudah kepada aras paling sukar secara visual (Bond & Fox, 2007). Proses kalibrasi instrumen ini melibatkan tiga peringkat. Pada peringkat pertama, anggaran parameter kesukaran item dan pada peringkat kedua pula adalah anggaran parameter keupayaan individu. Peringkat ketiga adalah anggaran parameter keupayaan penilai. Proses kalibrasi menunjukkan parameter item (kesukaran item), parameter individu dan parameter penilai dalam instrumen (kebolehan) dianggarkan supaya mereka boleh diletakkan dalam satu skala tunggal.

Kesukaran item, keupayaan individu dan kesan penilai dalam instrumen diukur menggunakan unit logit. Logit adalah log natural (dirujuk sebagai  $\ln$  atau  $e$ ) menghuraikan jumlah item yang betul berbanding jumlah yang tidak betul (Embretson & Reise, 2000; Siti Rahayah, 2013).

Pemetaan menggambarkan skala pengukuran daripada proses kalibrasi yang menunjukkan taburan kebolehan individu dan taburan kebolehan item. Individu yang berkebolehan tinggi terletak di sebelah atas kiri, sementara individu berkebolehan rendah terletak di sebelah bawah kiri. Item yang sukar dipersetujui terletak di sebelah



atas kanan, sementara item yang mudah dipersetujui terletak di sebelah bawah kanan. Manakala penilai yang tegas di sebelah atas tengah sementara penilai yang kurang tegas terletak di sebelah bawah tengah. Tujuan pemetaan adalah untuk menunjukkan hubungan antara kebolehan pelajar dengan aras kesukaran item (Bond & Fox, 2007) serta ditambah dengan peranan penilai. Pemetaan ini menunjukkan item yang sangat mudah dipersetujui (bahagian penilai juga bilangan responden yang berkebolehan tinggi (bahagian kiri atas peta) dan berkebolehan rendah (bahagian kiri bawah peta) (Siti Rahayah, 2008). Celahan pada kalangan taburan item turut dikaji bagi menguji sama ada terdapat item yang mencukupi untuk menguji semua kebolehan individu. Bagi pembentukan item instrumen, prinsip yang sama digunakan iaitu mengenal pasti individu yang mempunyai kebolehan berbeza dan mengenal pasti item dengan kesukaran yang berbeza.

**Model Pengukuran Rasch Pelbagai-Facet (PRPF).** Menurut Eckes (2005,2008), PRPF adalah model pengukuran lanjutan daripada Model Rasch yang boleh menggabungkan lebih *variable* (atau disebut *facet*) untuk mendapatkan hasil pengujian yang yang berhubung kait antaranya, iaitu contoh, analisis antara tiga facet iaitu, antara *examinees*, *raters*, dan *items*. Atau antara empat facet iaitu *examinees*, *raters*, *domains*, dan *items*. Facet bermaksud sebarang faktor atau pembolehubah, atau komponen dalam sesuatu pengukuran yang dijangka akan memberi kesan kepada bacaan skor dalam cara yang sistematik (Bachman, 2004).

Analisis pengujian PRPF adalah bertujuan untuk:

- a) Mengesan kebolehan individu melalui nilai yang positif (kebolehan tinggi) atau negatif (kebolehan rendah).
- b) Mengukur *severity* penilai.
- c) Mentaksir darjah konsistensi penilai
- d) Mengenal pasti potensi interaksi antara facet.
- e) Mengenal pasti kesukaran dalam facet yang dikaji.

Kajian Eckes (2005) menghuraikan prosedur untuk analisis yang sesuai menggunakan Model Pengukuran Pelbagai Faset adalah kebanyakannya bertujuan untuk: (a) mengukur *severity* (ketegasan) atau *leniency* penilai, kebolehan individu, kesukaran tugas (item) atau kriteria di dalam satu (b) memberikan skor yang adil kepada individu yang mengambil kira *severity* penilai, kesukaran tugas (item) (c) mentaksir darjah konsistensi penilai (d) mengesan pengaruh penilai yang lain (e) mengenal pasti kesan interaksi dan perbezaan fungsi facet. Kajian Barbot (2012) mendapati model PRPF membenarkan untuk mengenal pasti hubungan antara faktor-faktor dalam penulisan merentasi domain-domain dalam pada masa yang sama, mengawal interaksi antara ketegasan pakar dan penilaian bias antara pakar.

Penilaian terhadap prestasi kanak-kanak sukar untuk dijalankan secara objektif kerana ia bersifat subjektif. Pendekatan paling sesuai adalah menggunakan persetujuan antara penilai. Pelbagai pendekatan telah digunakan untuk mengenal pasti dan mengambil kira faktor persetujuan penilai dalam penilaian. Pendekatan yang dilihat sesuai adalah melalui latihan penilai dan mengawal skor yang diberi. Apakah ciri-ciri penilai dan item (tugas) dan aspek yang berkaitan yang sesuai semasa mentadbir instrumen penilaian prestasi? Bagaimana kebolehan kanak-kanak boleh ditentukan melalui penilaian prestasi ini? Adakah respon yang ditunjukkan terhadap item dalam

instrumen dapat menentukan ranking kebolehan kanak-kanak ini? Setelah item disahkan kebolehpercayaannya, adakah ia berupaya menggambarkan kebolehan kanak-kanak dengan skor yang adil? Jawapan terhadap persoalan ini terkandung dalam konsep dan prosedur model pengukuran pelbagai faset, iaitu teori dan metodologi pengukuran terkini yang relevan dengan penilaian prestasi seperti dalam situasi persoalan tadi.

Menurut McNamara (1996), peranan penting dalam teori pengukuran ini dipraktikkan dalam tiga perkara berikut:

- (1) menghuraikan dengan lebih jelas peranan penilai dan pengaruhnya dalam mentadbir instrumen berasas pentaksiran prestasi.
- (2) membolehkan kajian baru dijalankan dalam konteks baru,
- (3) menyesuaikan keadaan semasa mentadbir (faktor yang mempengaruhi) untuk membuat pengasingan terhadap pengaruh penilai dan pengaruh daripada faktor lain yang berkaitan dalam melaporkan prestasi peserta.

Model PRPF merupakan sejenis model pengukuran dan sesuai diguna untuk menguji instrumen. Model PRPF juga berupaya menukar data mentah kepada data sela (*interval*) yang sama nilai antara seunit dengan unit yang lain (Bond & Fox 2001). Model ini dapat memberi kefahaman kepada kita bagaimana kanak-kanak dan item 'berhubung' yang ditunjukkan serentak dalam satu analisis yang membentuk 'pembaris' kedudukan kebolehan kanak-kanak dan kesukaran item menggunakan skala yang sama.

Sebagaimana dalam CTT, model PRPF juga mempunyai andaian tertentu dalam pengukurannya. Pertama, mempunyai pelbagai kebolehan diukur oleh item-item yang membentuk sesuatu ujian (*unidimensional*). Kedua, mempunyai konsep setempat (*local independent*) iaitu ciri-ciri item tidak bergantung kepada kumpulan sampel dan kebolehan kanak-kanak tidak bergantung kepada ujian. Ketiga, fungsi ciri

item yang ditentukan menggambarkan perkaitan sebenar antara pemboleh ubah pendam (*latent trait*) atau kebolehan dengan pemboleh ubah nyata (respon murid). PRPF adalah pengukuran lanjutan yang dikenali sebagai Model Pengukuran Rasch Pelbagai Faset (Rasch, 1960, 1980).

Satu lagi aspek penting menggunakan model pengukuran Rasch Pelbagai Faset ini adalah selari dengan pembangunan instrumen yang menggunakan item berasaskan prestasi pada peringkat kanak-kanak yang menggunakan peranan penilai (*rater*) dalam proses pemerhatian. Modul PRPF berkebolehan mengesan pengaruh pemeriksa semasa memberi skor iaitu samada ketegasan atau kelembutan yang dikenali sebagai kekonsistenan pemeriksa. PRPF secara keseluruhannya mengambil kira interaksi *facets* (*person's ability, item difficulty & rater*). Kelebihan PRPF ini, parameter setiap *facets* tidak dipengaruhi oleh ciri-ciri sampel atau ujian. PRPF turut berkebolehan menguji kesan interaksi antara facets, yang dikenali sebagai analisis bias atau analisis interaksi. Setiap orang atau sampel adalah dipertimbangkan sebagai individu oleh PRPF dan penilaian kebolehan sampel dinilai secara berasingan (*independently*) berbanding facets yang lain (Linacre & Wright, 2002).

**Teori G.** Teori Generalizabiliti (GT) merupakan satu teori statistik berkaitan *dependability* bagi pengukuran perlakuan. GT dikemukakan oleh Cronbach, Rajaratnam & Gleser pada tahun 1963. GT mengemukakan alternatif yang fleksibel terhadap CTT. GT membenarkan pelbagai sumber ralat dalam pengukuran seperti item, tempat, dan pemeriksa dapat dianggarkan secara serentak termasuk interaksi antara sumber-sumber ralat tersebut (Shavelson & Web, 1991). GT ialah lanjutan kepada CTT, iaitu GT melanjutkan teori kebolehpercayaan dengan membenarkan pertimbangan multidimensi ke atas ralat penskoran yang seterusnya membawa kepada kerangka yang lebih tepat dan sistematik khususnya untuk penaksiran ujian pencapaian di mana GT melibatkan kepelbagaian kemahiran yang diuji (Brennan, 1992).

Menurut Shavelson & Webb (1991), sumbangan utama GT adalah dengan memberi penekanan kepada kepelbagaian sumber ralat dan mengurangkan peranan yang bersandar kepada rumusan pekali kebolehpercayaan mengikut CTT. GT menghasilkan dua koefisien kebolehpercayaan iaitu *koefisien G* dan koefisien Phi. *Koefisien G* dalam GT menyamai koefisien dalam CTT bagi rujukan norma atau keputusan *relative*. Dalam membuat keputusan *relative*, kepelbagaian ralat rawak ialah komponen varian bagi ralat dan semua komponen berinteraksi dengan responden. Koefisien Phi adalah koefisien kebolehpercayaan bagi rujukan kriteria atau keputusan mutlak. Bagi keputusan mutlak yang dilakukan, kepelbagaian ralat rawak jumlah semua sumber ralat tetapi tidak melibatkan komponen varians responden. GT mengemukakan dua jenis kajian iaitu kajian G (*G Study*) dan kajian D (*D Study*) iaitu kedua-duanya menyumbang fungsi tertentu dalam teori G.

Mengikut teori pengujian klasik setiap nilai skor yang didapati daripada sesuatu pengukuran mengandungi ralat. Skor yang diperolehi oleh seseorang responden dalam satu ujian mempunyai dua komponen iaitu ;

- i) skor sebenar (ukuran sebenar kebolehan responden - *true skor* )
- ii) skor ralat (disebabkan oleh alat dan kaedah pengukuran - *error score* )

Teori pengujian klasik merumuskan skor responden seperti berikut;

$$\mathbf{X} (\text{skor dicerap}) = \mathbf{T} (\text{skor sebenar}) + \mathbf{E} (\text{skor ralat})$$

Skor ralat dikira secara. Jumlah, tidak dapat dipastikan sumber-sumber yang banyak menyumbang kepada skor ralat tersebut.

Teori Generalizabiliti (GT) adalah lanjutan kepada teori ujian klasik (CTT) iaitu GT membenarkan anggaran magnitud pelbagai punca ralat secara serentak termasuk interaksi antara sumber-sumber tersebut ke atas ralat penskoran. Hal ini membolehkan pengiraan kebolehpercayaan yang lebih tepat ke atas skor responden. Melalui pengaplikasian GT, kebolehpercayaan sesuatu ujian tidak hanya bergantung pada satu

perkiraan tetapi sumber kepada ralat dikenal pasti dan kesannya kepada kebolehpercayaan ujian dapat diambil kira. Kepelbagaian aspek yang mempengaruhi dapatan seperti item, calon, interaksi antara item dan calon atau tempoh ujian yang boleh dikenal pasti secara terpisah mengikut GT. GT mampu untuk menilai dengan jelas sumber ralat dalam skor dapatan secara signifikan. Dengan cara ini guru dapat memperbaiki ujian yang ditadbir dan menangani masalah yang berkaitan dengan ralat penskoran dengan lebih sistematik. Hasil kajian dapat mengenal pasti sumber ralat dalam penilaian dan menganggarkan varians yang disumbangkan oleh setiap sumber ralat tersebut. Penggubal soalan boleh menganggarkan magnitud *relative* sumber ralat dan mengawal sumber ralat tersebut dalam penilaian yang akan datang, maka kebolehpercayaan ujian capaian yang ditadbir dapat ditingkatkan.

Antara konsep penting dalam GT ialah *universe of generalization* (UG). UG ialah ciri-ciri umum yang dikehendaki oleh penyelidik bagi membuat generalisasi dalam sebarang dapatan penyelidikan. UG boleh dijelaskan sebagai domain skor ujian yang dapat dirumuskan. UG mungkin terdiri daripada semua pemerhatian yang diterima dengan memasukkan *universe admissible of observation* (UAO) atau hanya subset kepada UAO. Definisi UG akan memberikan kesan kepada saiz koefisien generalizabiliti atau dengan perkataan lain anggaran kebolehpercayaan skor adalah UG. Keadaan ini memungkinkan penyelidik menganggarkan kebolehpercayaan skor ujian bagi tujuan tertentu yang mungkin berbeza. Brennan (1992) menyatakan bahawa ciri-ciri umum varian skor dan varian ralat adalah parameter penting dalam GT. Beliau menjelaskan bahawa ciri-ciri umum varian skor adalah sebagai “*variance for the object of measurement of their expected scores over all conditions of all facets in the universe of generalization*” (hlm.176). Ciri-ciri umum dan varian ralat adalah diungkapkan sebagai komponen varian skor atau komponen varian. Komponen varian adalah landasan pembentukan GT. GT dapat memisahkan sumber-sumber ralat atau

mengukur kesan setiap ralat dengan menghuraikan varian skor, atau ralat kepada komponen varian yang boleh dikenalpasti (Brennan & Johnson, 1995).

Shavelson & Webb (1991) menyatakan bahawa berbanding dengan skor ujian, skor sebenar ialah pengukuran ideal yang tidak mungkin diperolehi. Dalam GT, skor sebenar ialah skor purata bagi objek pengukuran (kebiasaanya individu) dengan mengambil kira semua kombinasi. Skor yang diperolehi mungkin boleh dibahagikan kepada beberapa kadar varian dari setiap keadaan pengukuran yang menjadi sumber ralat seperti item ujian, tempoh ujian, atau penilai. Sumber ralat ini dikenali sebagai *facet*. Menurut Lim (2008), pakar-pakar teori ini menjelaskan bagaimana skor diuraikan kepada pelbagai kesan faktor yang terlibat sehingga wujudnya komponen varians faktor berkenaan di bawah teori-G. Contohnya, bagi model kesan rawak satu hala *facet* tunggal responden ( $k \times i$ ), skor yang diperolehi oleh responden ke- $i$  dalam soalan (item) ke- $j$  diuraikan kepada sebutan min keseluruhan, kesan pelajar, kesan item dan kesan ralat. Daripada kesan responden, kesan item dan kesan ralat ini terhasilnya komponen varians faktor responden, faktor item dan faktor ralat masing-masing. Komponen varians faktor ini boleh diolah untuk membentuk fungsi komponen varians faktor yang dikenali sebagai pekali kebolehppercayaan (*G coefficient*) dalam teori-G. Setiap komponen varians faktor menggambarkan variasi faktor kajian. Misalnya, dalam suatu penilaian formal komponen varians faktor responden menunjukkan perbezaan kebolehan penguasaan pengetahuan atau kemahiran terhadap domain yang dinilai dalam kalangan responden. Komponen varians item mencerminkan darjah kesukaran antara satu item penilaian dengan item penilaian yang lain. Komponen varians ralat pula menggambarkan kewujudan ralat yang mempengaruhi kesahihan keseluruhan penilaian berkenaan. Umpamanya ralat boleh berpunca daripada beberapa faktor seperti faktor kesesuaian dan persekitaran penilaian

itu dijalankan, faktor kesihatan, faktor latihan penilai, faktor penilai atau faktor latar belakang responden.

Kajian ini akan mengenal pasti *dependability score*. *Dependability* merujuk kepada ketepatan dalam menggeneralisasikan skor yang diperoleh responden dalam sesuatu ujian kepada skor purata yang diperoleh oleh pelajar dalam pelbagai keadaan (Shavelson & Webb, 1991). Dalam konteks kajian ini, *dependability* ialah indeks yang diperoleh dalam analisis ujian berdasarkan perbezaan individu dan item yang dilihat dalam kajian G dan D.

Kajian G adalah mengenal pasti pelbagai sumber varians yang mungkin dalam penilaian dengan menganggarkan komponen varians yang disumbangkan oleh setiap satunya, kajian yang dijalankan untuk menilai kebergantungan pengukuran yang dilakukan kepada varians yang boleh dipertimbangkan dalam pengukuran yang akan datang (Shavelson & Web, 1991). Manakala Kajian D adalah mengemukakan pekali kebolehpercayaan sebagai pekali generalizabiliti merangkumi varians kepada sumber ralat. Kajian D juga dapat membezakan di antara keputusan *relative* dengan keputusan mutlak. Menggunakan maklumat yang telah diperoleh melalui kajian G, kajian D dapat mereka bentuk aplikasi pengukuran yang lebih baik dan sesuai untuk sesuatu pengukuran dan penaksiran yang dicadangkan (Shavelson & Web, 1991).

Secara ringkasnya kelebihan GT berbanding CTT ialah GT mempertimbangkan pelbagai sumber ralat secara serentak dan menjadi model yang lebih tepat dalam situasi pengukuran berbanding hanya kaedah satu model, sumber ralat dalam CTT yang hanya mempertimbangkan satu sumber ralat pengukuran bagi membuat keputusan yang *relative*. GT melengkapkan pendekatan secara bersepadu bagi menunjukkan kepelbagaian sumber ralat. Kaedah yang sama boleh diaplikasikan sama ada sumber ralat berkenaan ialah item, tempoh masa, bentuk atau



penilai. Tambahan pula GT dapat mempertimbangkan sebarang perbezaan sumber ralat dengan melibatkan kombinasi mana-mana facet berkenaan.

GT menghasilkan pendekatan yang lebih bersepadu bagi menaksirkan kebolehpercayaan yang dijalankan sama ada bagi tujuan membuat keputusan yang *relative* (ujian rujukan norma) atau keputusan sebenar (ujian rujukan kriteria). Keputusan yang *relative* adalah berdasarkan kedudukan individu dalam kumpulan berbanding keputusan sebenar. Keputusan sebenar adalah berdasarkan skor benar tanpa sebarang perbandingan dengan skor individu yang lain dalam kumpulan (Ary, Jacob & Razavieh, 1996). GT tidak membuat andaian tentang perbandingan sumber ralat tetapi menganggar secara serentak kepelbagaian sumber ralat termasuklah interaksi antara ralat-ralat berkenaan (Thomson, 1991). CTT mengandaikan bahawa wujud sumber secara bertindihan dan tidak mempertimbangkan kemungkinan bahawa ralat berkenaan menghasilkan ralat pengukuran yang lain atau ralat tambahan.

Menurut Crocker dan Algina (1986), penggunaan ANOVA digunakan untuk menganggarkan pekali kebolehpercayaan dan sumber ralat. Mengikut GT, *universe* merujuk kepada keseluruhan populasi keadaan pengukuran (*measurement condition*) dan populasi ini mengandungi pelbagai set keadaan pengukuran yang berkemungkinan. Setiap set ini akan menjadi *facet*.

Dalam contoh ini kemungkinan terdapat lima sumber utama kepada ralat iaitu;

- i. kanak-kanak (k)
- ii. interaksi antara penilai (*rater*) dengan kanak-kanak dan item (p:ki)
- iii. item (i)
- iv. interaksi kanak-kanak dan item (ki)
- v. *residual* (ralat yang tidak diketahui) (*e*)

skor yang diperhatikan adalah termasuk interaksi keempat-empat sumber ralat dibenarkan. Secara formulanya;

$$X_{pki} = \mu_k + \nu_p + \nu_i + \nu_{ki} + \nu_e$$

$X_{pki}$  = skor diperhatikan

$M$  = min populasi dan *universe*

$\nu$  = kesan ralat untuk interaksi yang dinyatakan

( $\nu_k, \nu_p, \nu_i, \nu_{ki}, \nu_e$ ) jelas dapat dilihat melalui kajian G, setiap sumber ralat yang dikenalpasti, dan dikira kesannya. Keadaan ini membenarkan pengiraan indeks kebolehpercayaan dalam sesuatu ujian atau pengukuran adalah lebih tepat dan lebih dipercayai.

**Analisis Melalui Perisian EduG.** EduG ialah satu perisian yang digunakan dalam teori Generalizabiliti. Perisian EduG dapat menganggar setiap komponen varian dan menentukan *dependability score* dalam sesuatu ujian. Terdapat berbagai reka bentuk boleh dianalisis menggunakan perisian EduG bergantung kepada facet yang ingin dikaji. Dalam kajian yang telah dijalankan, penyelidik telah menggunakan reka bentuk *Two-facet Partially Nested design*. Hasil analisis EduG dapat menghasilkan dua jenis kajian iaitu kajian G (*Generalizabiliti Studies*) dan kajian D (*Decision Studies*). Kajian G dapat mengenal pasti sumber varians dan magnitud varian manakala kajian D pula dapat menentukan *koefisien G* serta reka bentuk kesesuaian bilangan item dalam sesuatu ujian.

### **Model Pembinaan Instrumen**

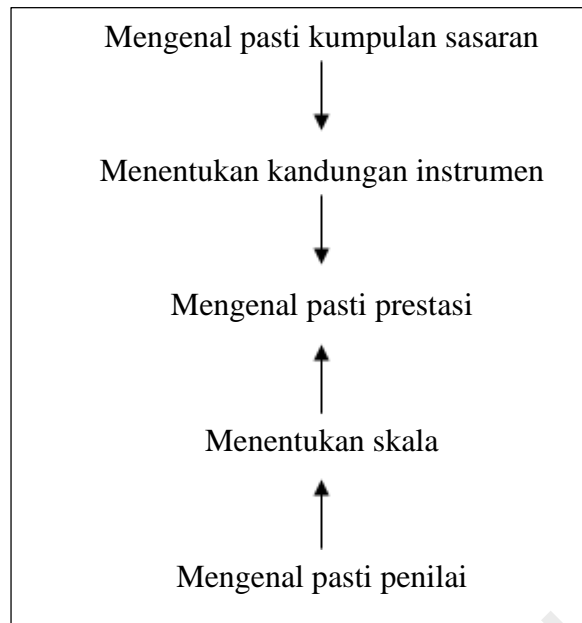
Peringkat pembinaan instrumen menggunakan model Miller ini bertepatan kerana langkah pembinaan instrumennya yang sesuai dengan model Rasch yang merupakan salah satu daripada Teori Respons Item (TRI). Proses pembinaan instrumen oleh Hambleton & Jones (1993) menunjukkan kesemua prosedur

pembinaan instrumen TRI dan CTT adalah sama kecuali perbezaan pada pengujian item pada lapangan, pembangunan instrumen dan analisis teknikal seperti kefungasian skala, unidimensionaliti, kajian kesahan dan kebolehpercayaan. Fasa rekabentuk dan pembinaan instrumen ini akan dibincangkan lanjut di dalam Bab 4.

Terdapat beberapa model pembinaan instrumen yang boleh digunakan untuk menghasilkan instrumen seperti dicadangkan oleh beberapa pengkaji lepas (Brown, 1983; Coaley, 2010; Cohen et al., 2012; Gregory, 2011; Hogan, 2007; Miller et al., 2013; Murphy & Davidshofer, 2005). Walaupun model - model ini mempunyai matlamat yang sama iaitu membina instrumen, namun proses pembinaannya adalah berbeza-beza. Setiap model mempunyai kelebihan yang tersendiri untuk digunakan.

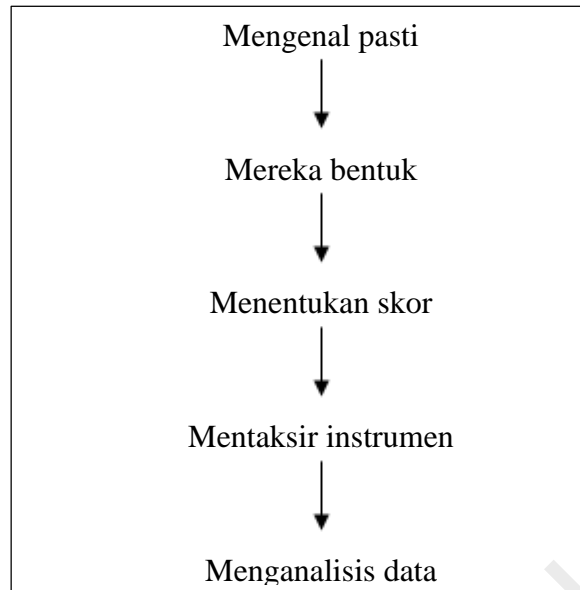
Model pembinaan instrumen yang digunakan bagi kajian ini adalah model yang dicadangkan Miller et al. (2013) seperti Rajah 3.1. Terdapat sepuluh proses dalam pembangunan instrumen iaitu (i) mentakrif konstruk, menentukan tujuan dan sasaran instrumen; (ii) mereka bentuk instrumen; (iii) membina item instrumen; (iv) menyediakan arahan pentadbiran dan penskoran instrumen; (v) mengendalikan kajian rintis; (vi) melakukan analisis item; (vii) menyemak instrumen; (viii) menguji kesahan dan kebolehpercayaan instrumen; (ix) menentukan norma dan (x) menyediakan manual. Berdasarkan proses dalam model ini, tiga peringkat telah ditetapkan untuk memastikan pembangunan instrumen lebih efektif iaitu peringkat reka bentuk, pembinaan dan pengesahan.

McNamara (1999) mengkhususkan dalam rekabentuk pembangunan instrumen berasaskan prestasi, terutamanya yang melibatkan pemilihan kandungan ujian dan format. Terdapat lima aspek yang ditekankan dalam merekabentuk pembangunan instrumen iaitu mengenal pasti kumpulan sasaran, menentukan kandungan instrumen, mengenal pasti prestasi, menentukan skala dan mengenal pasti penilai. Ringkasan rekabentuk instrumen menurut McNamara adalah seperti berikut:



Rajah 2.2. Proses Rekabentuk Instrumen oleh McNamara (1996)

Turner (2001) mengatakan instrumen yang dibina melalui kerangka pembinaan yang dirancang boleh mengukur kebolehan pelajar yang mempunyai nilai kesahan dan kebolehpercayaan. Peringkat pembinaan instrumen terdiri daripada lima langkah iaitu mengenal pasti, merekabentuk, menentukan skor, mentaksir instrumen dan menganalisis data. Mengenal pasti adalah perkara pertama yang ditentukan iaitu mengenal pasti standard yang digunakan di dalam instrumen. Merekabentuk merangkumi instrumen yang diadaptasi dan direka bentuk semula. Menentukan skor dengan merujuk kepada skala rating skor yang diberikan kepada tahap prestasi kanak-kanak. Mentaksir instrumen dilaksanakan berdasarkan objektif kajian, item yang ditaksir akan melalui penilaian oleh panel pakar dan kajian rintis dilaksanakan untuk mentaksir item yang telah disemak. Langkah terakhir adalah menganalisis data iaitu menggunakan prosedur tertentu untuk menganalisis data agar data yang dihasilkan mematuhi nilai kesahan dan kebolehpercayaan yang ditetapkan. Penyelidik kemudian membuat rekabentuk terakhir akan instrumen yang dibina. Ringkasan pembangunan instrumen menurut Turner adalah seperti Rajah 2.3 berikut:

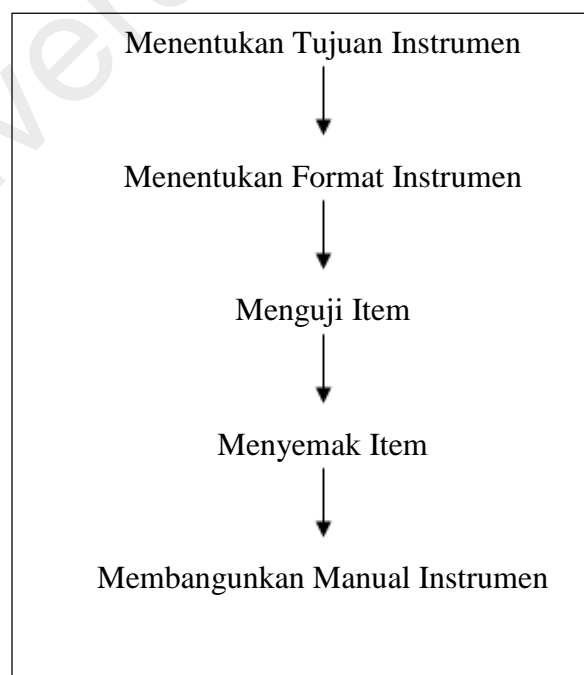


Rajah 2.3. Proses Pembangunan Instrumen oleh Turner (2001)

Menurut Wortham (2012), jika pengukuran dan penilaian kanak-kanak mengikuti kriteria untuk penilaian dalam abad baru, maka sistem penilaian perlu dibangunkan. Gabungan kaedah pengukuran yang digunakan akan bergantung kepada kegunaan untuk suatu sistem, tetapi secara keseluruhan, banyak komponen yang diperihalkan akan dimasukkan ke dalam mana-mana pelan untuk penilaian. Menggunakan sistem penilaian komprehensif melibatkan perancangan. Guru-guru bukan sahaja perlu memahami apa strategi dan alat-alat yang ada dan bagaimana untuk menggunakannya, tetapi mereka juga perlu mempunyai perancangan untuk menjalankan pentaksiran. Ujian standard direka untuk mengukur ciri-ciri individu. Ujian boleh diberikan kepada individu atau kumpulan. Tujuan ujian standard adalah untuk mengukur kebolehan, pencapaian, bakat, minat, sikap, nilai, dan ciri-ciri personaliti. Keputusan boleh digunakan untuk merancang pengajaran, mengkaji perbezaan di antara individu dan kumpulan, dan untuk bimbingan dan kaunseling.

Menurut beliau, terdapat lima langkah pembinaan instrumen standard yang perlu dilalui iaitu menentukan tujuan instrumen, menentukan format instrumen, menguji item, menyemak instrumen, dan membangunkan manual instrumen.

Menentukan tujuan adalah merangka pembentukan instrumen dan menentukan tujuan instrumen yang akan dibina. Menentukan format instrumen adalah prosedur yang ditentukan untuk penilaian kanak-kanak kecil dan tatacara pelaksanaan oleh pentadbir instrumen dimana ia mengandungi keseluruhan aspek untuk menjalankan penilaian berasaskan prestasi yang menggunakan tindak balas kanak-kanak terhadap tugas yang diberikan. Menyemak instrumen adalah memastikan kandungannya adalah kumpulan item yang telah diuji dan dianalisis, mengenal pasti item yang sukar dan ia akan ditadbir kepada lingkungan umur kanak-kanak yang berbeza. Item yang telah dimasukkan ke dalam instrumen akan disemak setelah borang penilaian digunakan dan ditadbir. Item yang dihimpunkan perlu untuk mengukur setiap domain perkembangan yang ditetapkan dan objektif kajian. Langkah terakhir dalam pembinaan instrumen adalah membangunkan manual instrumen. Tujuan manual dibangunkan untuk menerangkan tujuan instrument dibina, perkembangan instrumen dan prosedur piawai yang digunakan untuk mentadbir instrumen. Instrumen yang diterbitkan perlu mempunyai manual yang mesra pengguna. Ringkasan pembangunan instrumen menurut Wortham (2012) adalah seperti Rajah 2.4 berikut:



Rajah 2.4. Proses Pembinaan Instrumen oleh Wortham (2012)

Mematuhi piawai pembangunan instrumen yang berlandaskan *Standard for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME]* membantu dalam menyediakan kerangka yang sesuai untuk mengumpul dan melaporkan semua sumber bukti kesahan dan kesesuaian langkah tersebut. Terdapat sembilan langkah pembangunan instrumen seperti di Jadual 2.4 berikut:

Jadual 2.4  
*Sembilan Langkah Pembangunan Instrumen yang Efektif oleh Downing (2006)*

Langkah	Contoh Tugasan dalam Pembangunan Instrumen	Contoh Piawai Yang Berkaitan
1. Perancangan keseluruhan	Panduan sistematik bagi semua aktiviti pembangunan instrumen: konstruk; format instrumen; sumber utama bagi pembuktian kesahan, tujuan yang jelas; model psikometrik, jangka masa.	Piawai 1.1 Piawai 3.2 Piawai 3.9
2. Mendefinisikan kandungan	Plan persampelan; pelabagai kaedah yang berkaitan dengan tujuan penaksiran; sumber yang berkaitan dengan bukti kesahan kandungan; membentuk konstruk.	Piawai 1.6 Piawai 3.2 Piawai 3.11 Piawai 14.8
3. Spesifikasi instrumen	Definisi operasional; kerangka bagi bukti kesahan kandungan, ciri-ciri item yang diperlukan; rujukan norma atau kriteria.	Piawai 1.6 Piawai 3.2 Piawai 3.3 Piawai 3.4 Piawai 3.11
4. Rekabentuk instrumen	Membentuk format instrumen; memilih item untuk bentuk instrumen yang tertentu; kajian rintis	Piawai 3.7 Piawai 3.8
5. Penghasilan instrumen	Aktiviti penerbitan; percetakan; isu keselamatan; isu kesahan yang berkaitan dengan kawalan kualiti.	Piawai 3.18 Piawai 3.19
6. Mentadbir instrumen	Isu kesahan yang berkaitan dengan kepiawaian; isu keselamatan; isu masa.	Piawai 3.18 Piawai 3.19 Piawai 3.20 Piawai 3.21
7. Maklum balas pemarkahan instrumen	Isu kesahan; kawalan kualiti; kunci kesahan; analisis item	Piawai 3.6 Piawai 3.22
8. Melaporkan hasil instrumen	Isu kesahan; ketepatan; kawalan kualiti; jangkamasa; bermakna; isu –isu salah guna; cabaran; ambil semula.	Piawai 8.13 Piawai 11.12 Piawai 11.15 Piawai 13.19 Piawai 15.10 Piawai 11.11
9. Laporan teknikal instrumen	Sistematik; teliti; bukti kesahan didokumenkan secara terperinci; cadangan.	Piawai 3.1 Piawai 6.5

Fasa pembinaan instrumen ini dibina berdasarkan Model Miller et.al (2013) dengan gabungan langkah pembentukan instrumen oleh McNamara (1996), dan Turner (2001) dan disesuaikan dengan lapan langkah oleh Wortham (2012). Hasilnya adalah 10 langkah pembentukan instrumen bagi kajian ini. Penyelidik membahagikan langkah-langkah pembinaan instrumen ini kepada tiga peringkat, iaitu peringkat reka bentuk, pembinaan dan pengesahan agar kerangka pembinaan instrumen ini lebih sistematik. Pada peringkat reka bentuk, penyelidik membahagikan kepada 5 langkah iaitu, tujuan pembinaan instrumen, menentukan format penilaian, mengenal pasti kumpulan sasaran, menentukan kandungan instrumen, mengenal pasti prestasi, menentukan skala dan mengenal pasti penilai. Pada peringkat pembinaan, penyelidik membahagikannya kepada 5 langkah iaitu mengenal pasti standard, mengadaptasi instrumen (kajian rintis), mentaksir instrumen, menentukan tujuan pembinaan, dan format instrumen. Bagi peringkat pengesahan pula, terdapat 2 langkah iaitu kajian pengesahan dan analisis data. Menurut Cohen & Swerdlik (2005) proses pengesahan perlu diulang semula sekiranya analisis item yang dilaksanakan menunjukkan perlunya penambahbaikan item.

Bagi kajian ini, penyelidik mengulang proses pengesahan instrumen sekali lagi sepertimana yang disarankan oleh Kane (2006) bahawa perlu ada dua peringkat dalam proses pengesahan. Fasa pertama adalah peringkat pembinaan dan bertujuan untuk menghasilkan satu prosedur pengukuran yang munasabah. Apa yang kekurangan pada peringkat ini perlu dibetulkan sebaik mungkin sehingga instrumen yang dihasilkan adalah memuaskan. Bukti-bukti pengesahan pada fasa pertama bukan sahaja dari segi empirikal dan dokumentasi prosedur yang digunakan dalam pembinaan instrumen, tetapi juga berdasarkan daripada pandangan pakar yang dipilih. Dalam fasa kedua, penyelidik perlu melihat semula sama ada instrumen yang dibentuk berfungsi seperti yang dijangkakan dengan melakukan analisis pengesahan.



**Bukti-bukti Pengesanan.** Kesahan secara tradisional biasanya memerlukan pemeriksaan terhadap tiga jenis kesahan yang utama iaitu kesahan kandungan, kesahan kriteria dan kesahan konstruk. Wright & Stone (1999) menyatakan bahawa dengan apa yang dikatakan sebagai kesahan dengan cara tradisional kerana ketiga-tiga kesahan tersebut adalah berbeza dari segi maksud dan kaedah yang digunakan, sedangkan perkataan kesahan itu hanya mempunyai satu maksud yang dipersetujui oleh semua orang, namun begitu terdapat pelbagai kaedah yang perlu digunakan untuk menganalisis data yang ada. Menurut Wright & Stone (1999) satu cara untuk keluar daripada permasalahan ini adalah dengan memfokuskan pada data sedia ada atau respons sebenar individu terhadap item.

Kesahan sesuatu alat ukur menggambarkan sejauh mana ujian tersebut mengukur apa yang patut diukur (Aiken, 2000; Cohen & Swerdlik 2005; Gregory, 2007; Siti Rahayah, 2003). Definisi kesahan menurut *Standards for Educational and Psychological Testing* oleh *American Educational Research Education (AERA)*, *American Psychological Association (APA)* dan *National Council on Measurement in Education (NCME)* (1999) menyatakan bahawa sesuatu ujian mempunyai kesahan apabila inferens yang dibuat adalah sesuai, bermakna dan sesuai, bermakna dan berguna. Skor sesuatu ujian adalah tidak bermakna sehinggalah penguji membuat inferens terhadap skor tersebut berdasarkan manual ujian atau dapatan kajian yang lain (Gregory, 2007).

Menurut *Standard* yang dikeluarkan pada 1985 (APA 1974; AERA, APA & NCME 1985), terdapat tiga bentuk kesahan iaitu kesahan kandungan, kesahan kriteria dan kesahan konstruk yang relevan dengan pembangunan instrumen. Ketiga-tiga bentuk kesahan ini dikenali sebagai *trinatrian view* (Shepard, 1993). Pada tahun 1985, Standard telah mula menyatakan bahawa kesahan adalah satu konsep yang bersatu, namun begitu pembina ujian diberikan kebebasan untuk memilih standard untuk

kesahan (Shepard 1993). Dalam tahun 1999 Standards (AERA, APA & NCME 1999) mengatakan bahawa kesahan konstruk digambarkan sebagai satu konsep sepadu dan berada di bawah satu kerangka di mana ia mewakili semua bentuk kesahan. Semua jenis bukti kesahan, termasuk kesahan kandungan dan kesahan kriteria berada di bawah satu payung kesahan (Cohen & Swerdlik 2005).

Menurut Messick (1989), sebab utama kesahan dianggap sebagai satu konsep sepadu adalah kerana ia merupakan sumber yang komprehensif. Bukti kesahan konstruk biasanya diperoleh daripada integrasi bukti-bukti dari interpretasi skor ujian. Ia juga dianggap tergolong dalam kumpulan yang berkaitan dengan kesahan kandungan dan kriteria kerana maklumat mengenai kandungan kriteria diramalkan melalui skor ujian dan jelas menyumbang kepada interpretasi skor. Ini bermakna, pengukuran atau kesahan konstruk perlu dilakukan, bukan sahaja sebagai sokongan kepada interpretasi ujian tetapi juga untuk menjustifikasikan kegunaan ujian (Messick, 1988).

Berdasarkan Standard AERA, APA dan NCME (1999), semua skor ujian boleh dianggap sebagai indikator bagi sesetengah konstruk. Satu konstruk adalah konsep secara teori yang diperolehi daripada pelbagai bukti dalam pembinaan kerangka ujian. Hujah yang baik mengenai kesahan biasanya diperolehi daripada integrasi pelbagai sumber bukti yang koheren dengan teori untuk menyokong interpretasi skor ujian bagi kegunaan tertentu. Beberapa sumber bukti adalah seperti:

**Bukti berdasarkan kandungan instrumen.** Bukti daripada kandungan instrumen, termasuk analisis secara logik dan empirik yang mencukupi dengan kandungan instrumen bagi domain perkembangan dan domain perkembangan yang sesuai dengan interpretasi skor yang dicadangkan (AERA, APA & NCME 1999). Pengadil panel pakar boleh digunakan untuk menilai kepentingan, frekuensi dan kesesuaian item dengan domain perkembangan yang disenaraikan.

**Bukti berdasarkan proses respons.** Bukti daripada proses respon adalah berkaitan dengan proses yang melibatkan kesepadanan antara sifat dan konstruk penilai semasa membuat penilaian. Memandangkan kajian ini melibatkan penilai dalam membuat penilaian, bukti yang dikumpulkan bukan sahaja melibatkan penilaian terhadap kanak-kanak tetapi juga terhadap penggunaan skala pengukuran dengan menggunakan item-item dalam instrumen.

Sumber pertama yang dibincangkan dalam bukti berdasarkan proses respons adalah melibatkan kefungsiian skala pengukuran yang digunakan oleh para penilai dalam membuat penilaian. Penggunaan skala pengukuran lima katogeri skala pengukuran boleh menjadi satu proses yang rumit dan mengelirukan. Bagi memastikan penilai menggunakan skala pengukuran seperti mana yang sepatutnya, Penyelidik boleh melihat bagaimana skala pengukuran yang digunakan berfungsi dalam proses pengukuran yang dilaksanakan. Penyelidik boleh menilai kefungsiian skala pengukuran dengan menggunakan lapan garis panduan yang digariskan oleh Linacre (2004) yang penyelidik simpulkan kepada tiga bahagian iaitu pertama dengan memeriksa statistik kategori (iaitu frekuensi kategori dan min pengukuran) bagi setiap katogeri. (Andrich, 1978,1996; Bond & Fox 2001; Linacre, 1995, 1999). Yang kedua adalah dengan melihat nilai statistik kesepadanan (*fit*) statistik di mana ianya menilai kualiti skala pengukuran. Menurut Linacre (1999), nilai *outfit* MNSQ yang melebihi 2 menunjukkan terdapat “*more misinformation than information*”, yang bermaksud bagi kategori tertentu terdapat banyak keadaan yang tidak sesuai dalam proses pengukuran. Yang ketiga adalah dengan menilai ambang (*threshold*). Menurut Linacre (1999), nilai ambang perlu meningkat sekurang-kurangnya sebanyak 1.0 logit (bagi skala pengukuran lima kategori) untuk menunjukkan perbezaan jarak antara kategori yang

digunakan dan tidak melebihi 5 logit agar dapat mengelakkan jurang yang sangat besar di antara kategori yang digunakan.

Satu lagi analisis yang dapat membuktikan proses respons adalah melalui penilaian terhadap statistik kesepadanan (*fit*). Model pengukuran yang sesuai untuk menentukan kesahan ini adalah dengan menggunakan model Rasch. (Wright & Stones 1999). Penyelidik boleh menggunakan statistik kesepadanan untuk menentukan kesepadanan respons terhadap penilai dalam menilai pelajar, kesepadanan penilai dalam menggunakan items-item dalam IPHKT 1 dan 2. Statistik kesepadanan adalah indeks statistik yang menunjukkan sejauh manakah respons yang dilakukan oleh seseorang penilai itu sepadan dengan model pengukuran. Pendekatan yang biasa digunakan dalam menentukan data yang sepadan dengan model melibatkan pengiraan statistik kesepadanan yang menggunakan pendekatan khi-kuasadua Pearsonian (*Pearsonian chi-square*) (Wright & Panchapakesan, 1969).

Pendekatan ini menghasilkan dua jenis statistik kesepadanan iaitu *infit* dan *outfit*. Penyelidik biasanya melaporkan statistik kesepadanan ini dengan menggunakan dua bentuk iaitu statistik punca kuasadua (MNSQ) dan statistik pempaiwaan (ZSTD). Statistik outfit adalah berdasarkan pada jumlah kuasadua reja piawai (*standardized residuals*) dibahagi dengan darjah kebebasan. Nilai yang diharapkan adalah 1.0, dan nilai ini boleh berada di antara sifar hingga ke infiniti. Nilai yang besar 1.0 menunjukkan “kebisingan dalam data” dan nilai yang kurang dari 1.0 menunjukkan kurangnya kepelbagaian dalam data (Linacre & Wright 1994; Smith 2004; Wright & Masters 1982). Statistik *outfit* adalah sensitif terhadap *outlier*, terutamanya terhadap item yang mempunyai julat yang besar dalam kesukaran item dan keupayaan calon. Statistik *infit* pula adalah sensitif terhadap paten terhadap respons atau penilaian yang tidak dijangkakan. Secara matematikanya, *outfit* MNSQ adalah nilai statistik konvensional terhadap khi-kuasadua dibahagi dengan darjah kebebasan dan *infit*

MNSQ adalah *information-weighted variant* (varians maklumat pemberat) bagi *outfit* MNSQ. Nilai yang diharapkan bagi infit MNSQ juga adalah 1.0 dan boleh berada dalam julat sifat hingga ke infiniti (Linacre & Wright 1994; Smith 2004; Wright & Masters 1982). Dengan menggunakan nilai statistik kesepadanan, penyelidik boleh memeriksa kesepadanan respons bagi setiap penilai, pelajar dan item. Para penyelidik menganggap bahawa item-item yang mempunyai statistik kesepadanan pada julat yang sesuai sebagai item-item yang mempunyai kualiti yang tinggi.

**Bukti berdasarkan struktur dalaman.** Bukti yang berasaskan struktur dalaman adalah diperolehi daripada analisis struktur dalaman bagi sesuatu ujian. Menurut Downing (2003), struktur dalaman adalah ciri-ciri statistik atau psikometrik bagi sesuatu ujian dan item-item dalam ujian. Ini boleh ditunjukkan melalui darjah (*degree*) bagi hubungan antara item respon dan komponen domain perkembangan yang mematuhi konstruk berdasarkan skor ujian yang dicadangkan.

Menurut *Standard for Educational and Psychological Testing* (AERA, APA & NCME, 1999) terdapat beberapa bukti berdasarkan struktur dalaman yang boleh dijalankan bagi memeriksa struktur dalaman bagi sesuatu penilaian. Berdasarkan pendekatan model konstruk (Wilson 2004, 2005), pemeriksaan struktur dalaman bagi kajian ini adalah lebih kepada bukti kepada kawalan kualiti terhadap item-item. Bagi kajian ini, penyelidik menjalankan analisis psikometrik untuk menunjukkan item-item bagi instrumen ini menghasilkan satu pengukuran yang boleh dipercayai dan tepat respons yang diukur. Analisis psikometrik tersebut adalah (i) menganalisis kebolehpercayaan terhadap pengukuran yang dilaksanakan; (ii) menentukan sejauh mana kedudukan kesukaran item-item yang diukur sepadan dengan apa yang dijangkakan oleh model pengukuran dan (iii) menentukan tidak ada item (dan juga penilai serta kanak-kanak) yang berfungsi secara berbeza bagi subkumpulan umur yang berbeza, sekiranya mempunyai tahap prestasi yang sama.

Analisis kebolehpercayaan adalah sangat penting dalam kajian yang melibatkan struktur dalaman bagi sesuatu instrumen. *American Educational Research Association et. al.* (1999) menekankan tentang kepentingan mengumpul dan menilai bukti kepercayaan. Kebolehpercayaan merujuk kepada ketekalan pengukuran (Gregory, 2007; Mohd Majid, 1990; Nitko, 1983; Siti Rahayah, 2003). Kebolehpercayaan juga boleh dikatakan sebagai kestabilan, ketekalan dan boleh diulang semula (Walsh & Betz 1990). Menurut Nunnally (1978), kebolehpercayaan adalah ketekalan sesuatu ujian adalah hampir sama apabila sesuatu ukuran yang boleh dipercayai digunakan untuk mengukur tingkah laku yang sama pada masa berlainan. Alat ukur yang mengukur sesuatu pembolehubah dengan tekal dikatakan mempunyai kebolehpercayaan yang tinggi, sebaliknya alat ukur yang menghasilkan skor yang berubah-ubah bagi mengukur gagasan yang sama dikatakan tidak kekal dan kebolehpercayaan yang rendah. Menurut Nitko (1983), skor ujian mungkin tidak kekal kerana tingkah laku seseorang adalah tidak stabil, tidak dapat dijangkakan dari masa ke masa dari situasi ke situasi yang lain.

Bagi kajian ini, kebolehpercayaan yang dibincangkan adalah kebolehpercayaan konsistensi dalaman bagi kanak-kanak, item dan penilai. Perisian Facets yang digunakan bagi kajian ini membolehkan kajian terhadap kebolehpercayaan dilihat bagi ketiga-tiga faset yang digunakan iaitu kanak-kanak, item dan penilai. Indeks kebolehpercayaan dalam pengukuran yang digunakan dalam model pengukuran Rasch adalah kebolehpercayaan pengasingan (*separation reliability*) atau kebolehpercayaan Rasch. Nilai kebolehpercayaan Rasch bagi setiap faset menunjukkan sejauh mana berbezanya unsur-unsur dalam faset tersebut. Dalam lain perkataan, nilai kebolehpercayaan ini menunjukkan informasi mengenai sejauh mana unsur dalam sesuatu faset itu dapat dipisahkan untuk mendefinisikan kebolehpercayaan dalam sesuatu faset. (Engelhard, 2002; Myford & Wolfe 2003).

Kebolehpercayaan bagi faset kanak-kanak adalah seakan-akan sama dengan kebolehpercayaan dengan kaedah tradisional KR-20 atau nilai *Alpha Cronbach* yang biasa dilaporkan dalam literatur CTT. Nilai kebolehpercayaan adalah antara 0 – 1 dan nilai yang tinggi menunjukkan lebih pengasingan antara unsur-unsur dalam faset. Satu lagi indikator dalam konsistensi dalaman adalah nilai indeks pengasingan (*separation index*) yang menunjukkan taburan unsur-unsur pada faset tertentu dalam unit ralat piawai (Fisher, 1992). Bagi kajian ini, nilai kebolehpercayaan dan indeks pengasingan akan dilihat bagi faset item dan kanak-kanak sahaja. Bagi penilai pula, nilai kebolehpercayaan hasil daripada analisis Rasch, akan menunjukkan sejauh manakah ketegasan penilai berbeza di antara penilai dan ini berbeza daripada nilai kebolehpercayaan antara penilai yang ingin melihat bahawa penilai adalah mempunyai ketegasan yang sama. Memandangkan Model PRPF yang digunakan bagi kajian ini telah mengambil kira tentang ketegasan penilai bagi menghasilkan pengukuran yang adil kepada kanak-kanak, maka hasil analisis ini tidak akan dibincangkan. Bagi kajian ini, penyelidik menggunakan nilai kebolehpercayaan yang menunjukkan kebolehpercayaan penilai berdasarkan peratus persetujuan antara penilai.

Penyelidik juga mengumpulkan bukti keadilan dalam pengukuran dengan melihat analisis *bias* di antara faset bagi kajian ini. Analisis yang dijalankan adalah analisis DFF (*Differential Facets Functioning*). DFF boleh dilaksanakan terhadap faset penilai, item, kanak-kanak atau faset yang lain yang mungkin memberi kesan kepada ketidakadilan dalam pengukuran kanak-kanak dalam penilaian berasaskan prestasi. Menurut Angoff (1993), DIF (*Differential Item Functioning*) adalah merujuk kepada pemeriksaan terhadap item-item yang mempunyai ciri-ciri statistik yang berbeza terhadap seting kumpulan yang berbeza (setelah mengambil kira perbezaan kebolehan mengikut kriteria tertentu). Menurut Du et.al. (1996) definisi yang sama daripada segi prinsip boleh diaplikasikan bagi DFF. DFF boleh didefinisikan sebagai

terhadap item-item, penilai, kanak-kanak atau faset pengujian yang lain yang mempunyai ciri-ciri statistik yang berbeza terhadap seting kumpulan yang berbeza (setelah mengambil kira perbezaan kebolehan mengikut kriteria tertentu).

**Pentaksiran Kepelbagaian Pembelajaran Dan Perkembangan Kanak-Kanak Menggunakan Medium Pemerhatian.** Pentaksiran terhadap kanak-kanak dilakukan untuk mengenal pasti banyak perkara terutama untuk melihat tahap penerimaan pembelajaran dan proses perkembangannya. Medium utama untuk mentaksir kanak-kanak adalah melalui pemerhatian. Pemerhatian adalah kaedah yang biasa dan berkait terus dengan pembelajaran dan perkembangan kanak-kanak. Hal ini disebabkan ia memerlukan fokus ke atas setiap tingkah laku kanak-kanak dan melalui pemerhatian, kanak-kanak itu ditaksir dan keunikannya sebagai individu dapat diketahui (Wortham, 2012) dan di samping latihan yang meliputi kemahiran untuk memahami semua perkembangan kanak-kanak. Kemahiran memerhati adalah kemahiran penting untuk mentaksir kebolehan kanak-kanak agar objektif yang ditetapkan dapat dicapai dan maklumat yang dihasratkan dapat diperolehi (Billman & Sherman, 1997; Harrington, et.al., 1997; Pelo, 2006).

Terdapat tiga tujuan utama pemerhatian iaitu: (1) untuk memahami tingkah laku kanak-kanak, (2) untuk mentaksir perkembangan kanak-kanak, (3) untuk mentaksir perkembangan pembelajaran. Oleh kerana kanak-kanak masih belum menguasai bahasa dan kebolehan untuk membaca dan menulis, mereka tidak boleh mengekspresikan dengan jelas seperti orang dewasa. Mereka tidak boleh menunjukkan sejauhmana pengetahuan atau pemahaman melalui penilaian secara formal atau tidak formal yang melibatkan standard ujian dan tugas. Jalan terbaik untuk memahami tingkah laku kanak-kanak adalah dengan memerhati mereka dalam setiap aktiviti. Kanak-kanak berkomunikasi menggunakan bahasa badan. Cohen, Stern dan Balaban (1997) menyatakan, memerhati tingkah laku kanak-kanak memberikan maklumat



bagaimana fikiran dan perasaan kanak-kanak. Tingkah laku sosial kanak-kanak antara perkembangan yang sangat jelas dapat dilihat iaitu semasa interaksi sesama kanak-kanak, menjadi ahli kumpulan dan semasa melakukan aktiviti, sekaligus menunjukkan perkembangan sosial dan kemajuan tingkah laku.

Tujuan kedua pemerhatian adalah untuk mentaksir perkembangan kanak-kanak. Untuk tujuan ini, pemerhatian adalah bersifat spesifik. Pemerhati akan menentukan kemajuan dan kanak-kanak dalam perkembangan kognitif, fizikal, atau sosioemosi. Memerhati perkembangan kanak-kanak bukan sahaja untuk memahami turutan perkembangan kanak-kanak malah, mengetahui dan menyedari perkembangannya dan memberi bantuan jika berlaku kelewatan dalam domain perkembangan tertentu atau mengesan sebarang perkembangan lebih cepat.

Pentaksiran mengikut domain perkembangan dapat memberi pelbagai maklumat mengenai tahap perkembangan kanak-kanak dengan sistematik seperti yang dicadangkan oleh Beaty (2006) terhadap lapan perkara untuk melakukan dan merekodkan pemerhatian perkembangan kanak-kanak:

- (i) Untuk membuat penilaian awal terhadap kebolehan kanak-kanak.
- (ii) Untuk menentukan domain kekuatan kanak-kanak dan domain yang perlu dikuatkan.
- (iii) Untuk membuat perancangan individu berdasarkan keperluan yang dikesan.
- (iv) Untuk mengendalikan pemerhatian berterusan terhadap kemajuan kanak-kanak.
- (v) Untuk mempelajari dengan mendalam tentang perkembangan kanak-kanak dalam domain tertentu.
- (vi) Untuk menyelesaikan masalah tertentu yang melibatkan kanak-kanak.

(vii) Untuk melaporkan kepada ibu bapa atau pakar dalam perubahan, pertuturan dan kesihatan mental.

(viii) Untuk mengumpulkan maklumat dalam fail kanak-kanak dan digunakan untuk tujuan bimbingan atau penempatan.

Perkembangan kanak-kanak bersifat dalaman dan ia dapat ditaksir apabila dirangsang melalui tugas perkembangan dan aktiviti-aktiviti berasaskan ciri-ciri perkembangan kanak-kanak mengikut domain. Apabila kanak-kanak telah didedahkan dengan aktiviti pembelajaran, pentaksiran dilakukan untuk mengumpul maklumat terhadap pengetahuan kanak-kanak. Pemerhatian adalah alat dan kaedah yang sesuai untuk memahami gaya pembelajaran kanak-kanak. Tujuan ketiga pemerhatian adalah untuk mentaksir perkembangan pembelajaran. Aktiviti yang berkesan untuk memerhati perkembangan pembelajaran adalah melalui aktiviti berasas prestasi di mana kanak-kanak menunjukkan pembelajaran melalui performannya sebagai contoh, menunjukkan kemahiran motor di taman permainan. Aktiviti pembelajaran berasas prestasi boleh ditaksir dan dalam kajian ini, pentaksiran berasas aktiviti prestasi dijalankan untuk menentukan tahap pencapaian pembelajaran kanak-kanak.

Apabila berkait dengan pembelajaran, menggunakan Zon Perkembangan Proksimal (ZPP) oleh Vygotsky semasa proses pemerhatian dapat menentukan kemajuan kanak-kanak menguasai kemahiran (Bodrova & Leong, 2007). Vygotsky (1978) menyatakan, terdapat julat atau perbezaan zon tertentu antara apa yang kanak-kanak tidak boleh lakukan, boleh lakukan dengan bimbingan, dan boleh melakukan sendiri tanpa bimbingan. Pemerhatian yang dilakukan dapat menentukan di mana tahap perkembangan kanak-kanak di dalam ZPP. Sebagai contoh, pemerhatian dilakukan untuk menentukan tahap kecekapan kanak-kanak melukis, menggunting dan mengumpul objek selari dengan ZPP dalam aktiviti kemahiran motor halus yang diperlukan untuk tugas itu. Hal ini disebabkan setiap pemerhatian mesti ditafsir, dan

pemerhati perlu mengetahui bagaimana untuk menggunakan pemerhatian untuk mengumpulkan maklumat yang spesifik.

Latar belakang maklumat mengenai bagaimana kanak-kanak berkembang dan belajar sangat penting jika pemerhati ingin menterjemah tingkah laku kanak-kanak kepada maklumat yang boleh digunakan untuk memahami tahap perkembangan kanak-kanak untuk perancangan pada masa akan datang. Kualiti maklumat yang diperolehi daripada pemerhatian bergantung kepada kemahiran pemerhati. Wortham (2012) menegaskan pemerhati yang sofistikated berupaya menggunakan maklumat teori perkembangan dan tahap-tahap perkembangan kanak-kanak untuk mengenal pasti keadaan pemerhatian yang signifikan dan mentafsir keadaan ini dengan cara yang berguna dalam memahami kanak-kanak. Sebagai contoh, kanak-kanak yang bermain objek seperti butang dan mengelaskannya mengikut empat warna, dapat dikelaskan mengikut teori perkembangan kognitif Piaget berada pada peringkat 2-7 tahun yang berkebolehan membuat pengelasan.

Pemerhati perlu mempunyai sifat tertentu kerana proses pemerhatian adalah bukan memerhati semata-mata. Menurut Bentzen (1997) pemerhatian merupakan satu disiplin dan proses saintifik untuk mengesan tingkah laku menggunakan cara yang tertentu. Pemerhati mesti mengetahui apa yang ingin dilihat, bagaimana untuk merekod maklumat terkini dan bagaimana untuk menjelaskan tingkah laku tersebut. Kanak-kanak berkembang dengan cepat, dan tahap perkembangan mereka berubah secara berterusan. Melalui pemerhatian yang sistematik, perkembangan kanak-kanak dan tindak balas yang diberikan terhadap perubahan dan cabaran dapat dikesan dengan baik.

Menggunakan medium pemerhatian terhadap proses pembelajaran kanak-kanak dapat menyempurnakan amalan refleksi jika direkodkan. Refleksi pemerhatian berkait rapat dengan proses dokumentasi. Setiap maklumat yang diperolehi perlu

didokumentasikan dengan baik. Dokumentasi adalah strategi yang digunakan untuk mentaksir pembelajaran kanak-kanak. Pembelajaran di sekolah Reggio Emilia menggunakan dokumentasi untuk refleksi bagaimana kanak-kanak berhubung dalam pembelajaran mereka dan bukan untuk guru merancang untuk pengajaran. Cabaran yang dihadapi guru di sekolah Reggio Emilia adalah untuk melihat apakah sebenarnya berlaku dalam bilik darjah, untuk melihat interaksi, untuk mendengar perbualan sebenar, untuk merekodkan penerokaan dan untuk berkongsi dengan ibu bapa, kanak-kanak, dan guru-guru (MacDonald, 2006).

Dokumentasi boleh dikembangkan dengan pelbagai format untuk memenuhi kepelbagaian pembelajaran dan perkembangan kanak-kanak. Antara jenis dokumentasi adalah menggunakan portfolio individu (gambar, sampel hasil pembelajaran, sampel hasil kerja individu) dan pemerhatian terhadap perkembangan dan prestasi (*checklist*) dan boleh dikembangkan menggunakan web (Helm, Beneke, & Steinheimer, 2007; Seitz 2008).

**Pemerhatian Pembelajaran Dan Perkembangan.** Perkembangan kanak-kanak berkembang dengan cepat, berterusan dan berturutan. Pemerhatian perkembangan kanak-kanak perlu meliputi pelbagai domain perkembangan kerana menurut Bentzen (1997), semua kanak-kanak melalui peringkat perkembangan yang sama dan terdapat tiga ciri-ciri yang terlibat adalah, (1) tahap atau peringkat dalam perkembangan dan tidak berubah, kanak-kanak tidak melangkaui peringkat perkembangan; (2) kemajuan kanak-kanak melalui peringkat perkembangan adalah dalam kumpulan dan susunan yang sama; (3) Semua kanak-kanak, tidak mengira budaya atau perbezaan sosial, menunjukkan kemajuan pada peringkat perkembangan yang sama. Peringkat perkembangan adalah bersifat universal.

**Strategi Pemerhatian Perkembangan Kanak-Kanak.** Strategi pemerhatian terhadap perkembangan kanak-kanak yang ditekankan adalah terhadap kebolehan melalui setiap peringkat kebolehan. Walaupun kanak-kanak melalui peringkat perkembangan yang sama, tetapi kadar kebolehan dan keupayaan mereka melakukan tugas perkembangan adalah berbeza. Menurut Wortham (2012) perkembangan kanak dipengaruhi umur kronologi, kadar kematangan dan pengalaman individu. Terdapat kanak-kanak dalam kumpulan umur kronologi yang mempunyai perbezaan tahap perkembangan, yang mungkin disebabkan kadar kematangan mereka yang berbeza dan mempunyai pengalaman dan peluang yang berbeza. Kanak-kanak yang mempunyai peluang memanjat, berlari dan melompat dalam aktiviti luar berupaya menunjukkan kemahiran perkembangan motor yang lebih *advanced* berbanding kanak-kanak yang menghabiskan banyak masa bermain permainan berbentuk dalaman (*indoor games*).

Strategi pemerhatian melibatkan perkembangan fizikal, perkembangan sosioemosi, perkembangan kognitif dan perkembangan bahasa. Pemerhatian perkembangan fizikal memfokuskan perkembangan motor kasar dan motor halus. Kemahiran motor kasar melibatkan pergerakan dan kebolehan menggunakan otot-otot badan dalam aktiviti fizikal. Termasuklah pergerakan lokomotor seperti melompat, berlari dan memanjat, berguling, merangkak, melantun, menghayun, menderap (*galloping*) dan *skipping*. Semasa umur taska, sesetengah kanak-kanak sudah boleh mengayuh basikal dua roda dan menendang bola sepak (Johnson, 1998). Kemahiran motor halus melibatkan otot kecil badan seperti tangan dan jari. Kanak-kanak taska akan meningkatkan kawalan pergerakan jari, maka boleh melakukan aktiviti manipulatif seperti kemahiran makan, berpakaian, menulis, menggunakan permainan pembinaan dan melakukan tugas lain seperti belajar menggunakan *puzzle*, menggunakan gunting, menggunakan berus, pen, pensil, blok-blok kecil dan tanah liat.

Pemerhatian perkembangan sosioemosi adalah domain yang signifikan dengan perkembangan kanak-kanak taska. Pada tempoh ini, kanak-kanak mula bertukar daripada sifat egosentrik kepada interaksi sosial dengan orang lain. Perkembangan emosi adalah selari dengan perkembangan sosial. Kanak-kanak taska mempunyai sifat yang halus dan emosi seperti tidak gembira, marah, seronok, cemburu dan takut. Menurut Bentzen (1997) emosi yang paling biasa dikenali di taska adalah agresif, suka bergantung dan takut. Agresif adalah kelakuan yang boleh mengganggu orang lain atau peralatan di kelas. Suka bergantung disebabkan mencari perhatian, bantuan, penerimaan atau mendapatkan ketenangan. Manakala takut adalah termasuk kelakuan suka menangis dan mengelakkan situasi yang menakutkan.

Menurut Berger (2000) & Berk (2001), ciri-ciri penting perkembangan sosioemosi adalah *self-concept*, *self esteem* dan *self-regulation*. Melalui *Self-concept*, kanak-kanak mempunyai kesedaran tentang perbezaannya dengan kanak-kanak lain dan mempunyai ciri individu iaitu dengan penguasaan kemahiran dan kecekapan tertentu. *Self esteem* hadir apabila kanak-kanak mula membuat penilaian tentang diri dan kebolehannya. *Self-regulation* terhasil apabila kanak-kanak mengembangkan kesedaran terhadap perasaan mereka dan mula bertingkah laku yang membolehkan mereka mengawal diri.

Perkembangan kognitif berkaitan dengan fungsi mental yang berkaitan dengan bagaimana kanak-kanak belajar dan memahami tentang dunia. Pemerhatian terhadap perkembangan kognitif kanak-kanak taska bermula dengan tahap praoperasi iaitu kebolehan untuk menggunakan bahasa telah berkembang. Bermula dengan bahasa, mental mula memproses untuk memahami maklumat yang diperolehi dan dibina. Piaget menyifatkan perkembangan kognitif berkaitan dengan kematangan, pengalaman, dan proses sosial. Selain itu, keluarga kanak-kanak, persekitaran dan peluang kepada pengalaman memberi kesan kepada perkembangan kebolehan kognitif

kanak-kanak. Pengetahuan akan disusun atur sebagaimana kanak-kanak menyusun pengalaman yang dilaluinya untuk meneliti dan mengembangkan pemahamannya.

Perkembangan bahasa berkait dengan perkembangan kognitif. Semasa umur taska, kanak-kanak mempelajari lebih kurang 10,000 perkataan (Wortham, 2012). Mereka mula mempelajari peraturan menggunakan bahasa iaitu morfologi, sistaksis dan semantik. Morfologi dan sistaksis berkaitan dengan kefahaman terhadap bunyi dan tatabahasa, manakala semantik menjelaskan perbendahaharaan kata dan makna. Pemerhatian perkembangan bahasa melibatkan kebolehan mereka menyertai perbualan dan berkomunikasi. Selain itu, mengetahui peraturan untuk berkomunikasi iaitu pada waktu yang sesuai untuk menggunakan perkataan dan bahasa yang sesuai. Apabila mencapai umur 4 tahun, kanak-kanak telah memahami untuk mengendalikan perbualan mengikut bahasa dan budaya mereka. Oleh yang demikian, kepentingan pemerhatian perkembangan bahasa untuk peringkat taska ini adalah, untuk mengetahui kebolehan kanak-kanak berkomunikasi, memahami perbezaan antara percakapan secara egosentrik atau sosial dan mengetahui bagaimana kanak-kanak menggunakan sintaksis, tatabahasa dan perbendahaharaan kata dalam proses mengembangkan bahasa mereka.

**Kriteria Pembuktian Kepelbagaian Pembelajaran Perkembangan Kanak-Kanak Berdasarkan Prinsip Pentaksiran Taman Asuhan Kanak-Kanak.**

Perbezaan pentaksiran kanak-kanak dengan kumpulan umur yang lain adalah merujuk bagaimana pelaksanaannya. Seorang guru hendak melihat kemahiran, pengetahuan dan sikap kanak-kanak secara menyeluruh (holistik) semasa pemerhatian dijalankan. Pentaksiran dijalankan bukan semata-mata untuk melihat pencapaian yang selari dengan kurikulum kebangsaan, tetapi lebih menekankan bukti atau hasil kerja yang memperlihatkan pencapaian. Kanak-kanak adalah istimewa dan tersendiri. Geva & Kelly (1992) berpendapat perkembangan yang berbeza antara kanak-kanak

menyebabkan pentaksiran yang digunakan perlu memfokuskan ciri-ciri istimewa pada kanak-kanak dan semasa melibatkan dalam aktiviti secara spontan.

Beliau membezakan aspek perkembangan kanak-kanak yang meliputi ciri-ciri pada sikap, perasaan, sosial dan fizikal dengan memerlukan pentaksiran yang mengambilkira keseluruhan perkembangan ini bagi setiap kanak-kanak secara individu. Kriteria pertama, wujudnya aspek perkembangan yang berkaitan satu sama lain, maka, semasa pentaksiran terhadap seseorang kanak-kanak aspek perkembangan perlu dihubungkan agar dapat mengetahui kanak-kanak ini menunjukkan perkembangan normal, *advance* atau *delay*. Sebagai contoh, seorang kanak-kanak yang mempunyai masalah penumpuan dilihat sempurna dari segi fizikal, namun tidak matang dari segi komunikasi dan sosial hingga mengganggu penerimaan sebarang bentuk arahan. Maka pentaksiran yang baik perlu membuktikan semua perkembangan kanak-kanak konsisten satu sama lain. Jika berlaku sebaliknya, ini membuktikan pencapaian kanak-kanak dipengaruhi dengan kesukaran tugas yang diberikan atau faktor lain. Maka tindakan seterusnya boleh diambil.

Kriteria kedua ialah wujudnya pembelajaran yang tidak boleh diletak dibawah satu matapelajaran dalam konteks kanak-kanak. Geva & Kelly (1992) menyatakan semasa aktiviti memasak, pertambahan pengalaman pembelajaran dalam bahasa (perbincangan resepi dan proses), matematik (semasa mengira sudu, membanding saiz, berat), sains (semasa pemerhatian proses perubahan), rekabentuk dan teknologi (memerhatikan peranan peralatan memasak yang digunakan), kesihatan (semasa membincangkan aspek kebersihan menyediakan makanan). Penerimaan pengalaman ini berbeza bagi setiap kanak-kanak. Oleh yang demikian, semasa proses pemerhatian dan pentaksiran hendaklah dijalankan terhadap kepelbagaian pengalaman pembelajaran di bawah satu aktiviti. Setiap item dalam pentaksiran perlu menyediakan



peluang kepada kanak-kanak untuk menunjukkan pencapaian yang merentasi keseluruhan kurikulum pembelajaran.

Kriteria ketiga ialah wujudnya pencapaian yang berbeza sekiranya kanak-kanak diletakkan dalam situasi yang mereka tidak biasa. Pembelajaran yang terbaik untuk kanak-kanak adalah dalam persekitaran taska dan guru mereka. Kanak-kanak lebih termotivasi untuk melakukan aktiviti tersebut. Pentaksiran yang dijalankan secara pemerhatian dalam situasi pembelajaran semulajadi dapat mencatatkan pencapaian kanak-kanak dengan lebih baik. Kriteria keempat ialah wujudnya keperluan pihak taska untuk menyediakan persekitaran agar membangunkan kanak-kanak memiliki autonomi dan mendisiplinkan diri. Prosedur pentaksiran perlu memberi kesan kepada dua keperluan ini, dan meletakkan kanak-kanak dalam situasi yang berdikari di mana mereka boleh diharapkan menjadi lebih matang dan membuat tugas yang lebih sukar. Situasi ini boleh dilihat semasa pentaksiran ditadbir. Tugas yang diberikan akan menentukan sama ada, kanak-kanak ini perlukan bimbingan atau sebaliknya. Pada peringkat ini, proses pembelajaran dilihat lebih penting berbanding hasil yang ditunjukkan. Kebolehan menjadi kreatif, tabah, individu bermaklumat dan penyelesaian masalah adalah lebih signifikan. Kebolehan setiap kanak-kanak adalah berbeza, maka pentaksiran perlu membuktikan kepelbagaian kebolehan kanak-kanak dalam aspek perkembangan.

Kriteria kelima ialah wujudnya sokongan pendidik dan pengamal pendidikan awal kanak-kanak yang mempercayai permulaan seseorang kanak-kanak belajar adalah mengetahui apakah yang mereka boleh lakukan dan bukan apa yang mereka tidak boleh lakukan. Prosedur pentaksiran perlu berupaya menentukan apa yang kanak-kanak boleh lakukan dan persediaan yang boleh dilakukan kepada setiap kanak-kanak agar pembelajaran dapat dilanjutkan. Oleh yang demikian, pelbagai tugas disediakan agar memenuhi kepelbagaian tahap kebolehan kanak-kanak. Maka hasil

pentaksiran perlu mengenal pasti kebolehan kanak-kanak dan kesukaran tugas agar dapat menentukan kanak-kanak yang boleh melakukan dengan baik dan menentukan tahap kesukaran tugas agar kanak-kanak masih berpeluang untuk mencapainya. Pada peringkat ini, bimbingan guru diperlukan agar kanak-kanak dapat melakukan tugas dengan lebih cekap. Pemilihan pendekatan untuk menganalisa data juga memainkan peranan penting agar ia memberi skor yang adil kepada kanak-kanak mengikut ranking kebolehannya.

Kriteria keenam ialah wujudnya faktor daripada hubungan rakan dan orang dewasa yang memberi kesan kepada pembelajaran dan pencapaian kanak-kanak. Hubungan ini berupaya meningkatkan harga diri dan pembelajaran kanak-kanak atau menggagalkan keyakinan mereka dan melemahkan kebolehan untuk belajar. Faktor interpersonal ini tidak boleh diabaikan semasa mentaksir kanak-kanak. Pentaksiran boleh mengambil kira faktor persekitaran terdekat yang berhubungan dengan kanak-kanak untuk mentaksir kepelbagaian perkembangan yang dipengaruhinya.

**Pentaksiran Berasaskan Prestasi.** Pentaksiran berasas prestasi memerlukan lebih maklumat semasa mengukur apa yang kanak-kanak boleh lakukan atau praktikkan, sebagai tambahan kepada apa yang kanak-kanak tahu (Pierson & Beck, 1993; Wiggins, 1993). Lebih lagi, pentaksiran berasas prestasi termasuk melengkapkan tugas dalam konteks sebenar. Pentaksiran berbentuk ini adalah alternatif yang positif untuk mengukur pembelajaran dan perkembangan kanak-kanak berbeza pendekatan tradisional seperti ujian pencapaian, ujian kecerdasan yang mengukur pembelajaran kanak-kanak dengan memfokuskan apa yang kanak-kanak tahu khususnya kognitif sahaja.

Pembinaan instrumen dalam kajian ini berasaskan prestasi kerana semua domain dalam instrumen ini mempunyai item-item berbentuk prestasi, iaitu domain kognitif, bahasa, fizikal, kreativiti sosioemosi dan rohani. Di sini perlu dibincangkan

pentaksiran prestasi daripada beberapa aspek, termasuklah definisi, keperluan dan kepentingannya, prinsip pembinaan instrumen pentaksiran prestasi, jenis-jenis pentaksiran prestasi, dan juga teknik penskoran skala likert.

***Definisi Penilaian Berasas Prestasi.*** Mengikut Mohamad Sahari (2002), menilai prestasi bukanlah satu perkara baru. Tambahnya lagi, sejak berabad lamanya kecekapan seperti berenang dan memainkan alat muzik telah pun mempunyai cara-cara tersendiri untuk dinilai. Sungguhpun begitu, penilaian prestasi yang difahami sekarang adalah perkembangan baru dalam sistem pendidikan Negara. Sebelum perbincangan mengenai keperluan, kepentingan serta syarat-syarat penilaian prestasi dalam pendidikan diteruskan, diutarakan beberapa definisi penilaian prestasi mengikut tokoh dalam bidang ini:

Penilaian berasaskan prestasi mengikut Hibbard (1996) adalah “mempersembahkan satu set strategi untuk ..... mengaplikasikan pengetahuan, kemahiran, dan tabiat kerja melalui tugas-tugas prestasi yang melibatkan pelajar secara bermakna”

Mengikut Nitko (2003), penilaian prestasi didefinisikan sama dengan pentaksiran alternatif (*alternative assessment*) dan saling bertukar penggunaan mengikut konteks tertentu. Pentaksiran alternatif adalah suatu tugas yang memerlukan murid menunjukkan pencapaian melalui hasil keluaran dalam bentuk tulisan dan percakapan sama ada secara berkumpulan atau secara individu. Menurut Nitko (2001) lagi, penilaian prestasi memerlukan murid menjalankan kerja secara ‘*hands on*’, aktiviti-aktiviti yang memerlukan aplikasi pengetahuan dan kemahiran dengan menggunakan kriteria yang jelas. Definisi Nitko diperjelaskan lagi oleh Mohamad Sahari (2002), di mana beliau mendefinisikan pentaksiran prestasi sebagai “proses pengukuran dan penilaian yang meminta murid mencipta satu jawapan atau produk untuk menampilkan pengetahuan atau kemahiran yang dipelajari olehnya”.

Mohamad Sahari (2002) menjelaskan lagi, semua bentuk tingkah laku, baik yang berkaitan dengan kemahiran kognitif yang kompleks sehinggalah cara murid bertingkah laku dalam situasi sosial adalah merangkumi konsep pentaksiran prestasi. Mohamad Sahari kemudiannya mendefinisi pentaksiran prestasi sebagai usaha sistematik dan berterusan untuk mengukur dan menilai pencapaian proses atau produk murid, sama ada berbentuk konsep atau kemahiran yang rumit dan kompleks.

Pencapaian diukur menggunakan pemerhatian dan penilaian profesional, dan bentuk pengukuran ini dipanggil penilaian berasas prestasi (Stiggins, 1987). Terdapat perbezaan ciri-ciri penilaian tradisional dan berasas prestasi. Berdasarkan penilaian tradisional, skor diberikan terus daripada instrumen penilaian, iaitu setiap item menawarkan kepada calon nombor sebagai aneka pilihan dan hanya satu pilihan adalah tepat, manakala penilaian berasas prestasi mencungkil atau memerlukan respon yang akan diadili oleh penilai. Penilaian berasaskan prestasi memerlukan penilaian yang subjektif (McNamara, 1996). Pentaksiran prestasi juga dikenali dalam istilah lain sebagai pentaksiran autentik atau prestasi autentik (Wortham, 2012).

***Pembelajaran Autentik Dan Pentaksiran Autentik.*** Menyelesaikan tugas dalam konteks yang sebenar adalah termasuk dalam pentaksiran berasas prestasi. Menyelesaikan tugas akan mempamerkan kebolehan sebenar dalam diri kanak-kanak. Apabila menyebut kebolehan sebenar ini, ia selari dengan istilah autentik iaitu pentaksiran autentik (pentaksiran sebenar) atau prestasi autentik (prestasi sebenar). Jika menggunakan pentaksiran prestasi atau autentik untuk memahami bagaimana kanak-kanak boleh mengaitkan atau menggunakan apa yang telah mereka pelajari, pengalaman pembelajaran yang disediakan juga mestilah autentik dan bermakna. Apabila kanak-kanak berhubung dengan pembelajaran yang autentik, mereka diberi peluang untuk menghubungkan maklumat baru dengan maklumat yang sedia ada semasa menyelesaikan masalah.

Menjelaskan lagi perkaitan pembelajaran autentik dengan kebolehan sebenar kanak-kanak adalah merujuk kepada Kleinert (2002) yang menyatakan tujuan pendekatan ini adalah untuk membolehkan kanak-kanak menunjukkan bagaimana mereka gunakan apa yang mereka faham dan untuk mempersembahkan pembelajaran dalam bentuk produk atau prestasi. Ini bermaksud, melalui pembelajaran autentik, ia telah merangsang kanak-kanak mempamerkan pengetahuan yang dimiliki atau perasaan sebenar dalam diri mereka. Menurut Wehlage (1996), pembelajaran autentik berdasarkan pembinaan pengetahuan dan berfokus kepada pemikiran aras tinggi. Tujuannya adalah untuk meningkatkan tahap pengetahuan dan membina pengetahuan baru. Maka, bagi kanak-kanak, peluang yang diberikan melalui pelbagai aktiviti semasa pentaksiran adalah bertujuan untuk melihat tahap pengetahuan yang dimiliki dan dapat pengalaman baru serta ilmu yang baru apabila bantuan diberikan semasa melakukan aktiviti.

***Tujuan pentaksiran prestasi.*** Ujian tradisional seperti ujian pencapaian, ujian persediaan dan ujian kecerdasan selalunya direka bentuk untuk menapis (*sort*) (Mohamad Sahari, 2002) dan menentukan kedudukan (*ranking*) berdasarkan skor tunggal (*single score*). Lebih buruk lagi kesan ujian tradisional adalah untuk menyingkir murid. Sebagaimana menurut Michael (1991) dan Conaghan (1994) dalam Rohizani (2003), kelemahan pentaksiran seperti apa yang berlaku dalam peperiksaan sekarang adalah lebih merupakan satu terminal untuk 'menyingkirkan' murid yang tidak dapat mencapai satu piawai yang telah ditetapkan oleh sesuatu peperiksaan awam. Apabila percaya bahawa tujuan pentaksiran untuk menapis, sebagai wadah menentukan ranking dan menyingkir murid, menyebabkan ahli-ahli psikologi menggunakan banyak masa untuk menetapkan kedudukan murid daripada membantu mereka (Gardner 1993).

Tujuan pentaksiran pembelajaran perkembangan berasas prestasi menggunakan instrumen adalah untuk menyokong kanak-kanak melalui eviden dan bukti yang diperolehi serta mengenal pasti kekuatan dan kelemahan (Gardner, 1993). Apabila merujuk tujuan pertama menggunakan pentaksiran berasas prestasi kepada kanak-kanak, pentaksiran ini merupakan alat yang bagus untuk mentaksir kemajuan dalam perkembangan kanak-kanak kerana pentaksiran berasas prestasi telah direka bentuk untuk mengukur prestasi kanak-kanak yang sebenar atau mereka tugas atau aktiviti yang berkaitan untuk pembelajaran. Menggunakan pemerhatian kepada prestasi yang berkait rapat atau berhubung terus kepada perkembangan kadar pencapaian kanak-kanak (Harrington et al., 1997). Kedua, pentaksiran berasas prestasi adalah diintegrasikan kepada pengajaran. aktiviti prestasi adalah hasil pembelajaran semulajadi selari dengan kurikulum dan pengajaran dan tidak dipisahkan.

Hills (1993) menghuraikan,

“ Apabila menggunakan pentaksiran berasas prestasi, guru perlu mengetahui bagaimana kesesuaian rekabentuknya, hubungan sebagai alat pengujian, mentafsir hasil pentaksiran untuk memahami kemajuan kanak-kanak dan merancang pengajaran yang selanjutnya dan menyampaikan hasil pentaksiran prestasi kepada ibu bapa dan pentadbir.”

Seterusnya, pentaksiran berasas prestasi digunakan untuk mentaksir program TASKA memenuhi keperluan kanak-kanak. Pentaksiran yang baik dapat menjelaskan sasaran objektif TASKA untuk menyediakan kurikulum berkaitan perkembangan. Interaksi.

***Keperluan dan Kepentingan Pentaksiran Prestasi.*** Dalam jangka pendek, pentaksiran berasaskan prestasi menghasilkan maklumat berguna berkenaan apa murid dapat buat. Dalam jangka panjang, murid akan menjadi seorang warganegara yang baik, pekerja produktif dan semuanya mencapai matlamat yang dikehendaki (Brand,

1997). Pendapat ini disokong oleh Wigdor & Gardner (1982) dalam Torff (1997) dengan menyatakan berapa baik sekalipun ujian piawai atau ujian kebolehan kognitif, puratanya hanya menyumbang empat peratus daripada varian dalam prestasi kerja, 96% lagi tidak ditunjukkan dalam ujian-ujian berkenaan. Gardner (Checkley 1997:12) menyatakan:

“Penekanan ke atas pentaksiran prestasi sekarang sangat disokong kuat oleh teori kecerdasan pelbagai...pertamanya, jangan melihat ujian jawapan pendek sebagai sesuatu ‘tapisan’ sahaja, tetapi hendaklah menilai prestasinya sama ada bidang linguistik, logikal, estetika, atau sosial, dan kedua, jangan berpendirian bahawa pentaksiran kami hanya untuk mengukur sesuatu sahaja, tetapi pentaksiran yang selalu menjuruskan kepada murid menunjukkan kefahaman dalam pelbagai cara”.

Nitko (2001), mempercayai bahawa pentaksiran prestasi mempunyai potensi untuk mengundang murid menunjukkan pencapaian dan boleh dinilai dalam pelbagai cara yang berbeza.

Kelebihan pentaksiran prestasi mengikut Nitko (2001,2003), antaranya dapat mencapai matlamat pembelajaran yang kompleks, mentaksir kebolehan ‘untuk membuat’, di mana dapat melihat murid menyelesaikan masalah dalam kehidupan sebenar, pentaksiran menjadi lebih konsisten dengan teori pembelajaran moden. Menurut Linn & Gronlund (1995), teori pembelajaran moden mengesyorkan murid aktif terlibat dalam membentuk sesuatu makna. Tambah Nitko (2001) lagi, pentaksiran prestasi dapat mengintegrasikan pengetahuan, kemahiran dan kebolehan, boleh diselitkan dalam aktiviti pengajaran, meluaskan pentaksiran murid dan menggalakkan guru-guru mentaksir proses murid melaksanakan sesuatu tugas, berbanding mentaksir produk yang dihasilkan. Pentaksiran proses dan prosedur membuat sesuatu,

mendedahkan kekuatan dan kelemahan murid serta boleh mengetahui teknik dan strategi melakukannya.

Pentaksiran prestasi juga boleh mentaksir hasil pembelajaran yang kompleks dan menyeluruh di mana ianya tidak dapat dilakukan oleh kaedah yang lain (Mohamad Sahari, 2002). Kaedah lain yang dimaksudkan di sini adalah kaedah ujian piawai atau ujian kertas dan pensel yang mempunyai kelemahan tertentu. Keperluan kepada pentaksiran prestasi adalah bertitik tolak daripada kelemahan-kelemahan ujian piawai atau ujian-ujian rujukan norma yang dapat dikesan.

Nitko (2001), mempercayai bahawa pentaksiran prestasi dalam kecerdasan pelbagai domain mempunyai potensi untuk mengundang murid menunjukkan pencapaian dan boleh dinilai dalam pelbagai cara yang berbeza. Amy (2000) berpendapat taksiran berasaskan prestasi memerlukan individu menggunakan pengetahuan dan kemahiran dalam 'konteks', tidak semata-mata melengkapkan tugas sebagai syarat sahaja. Mengikut Gardner (1993), kebanyakan pihak yang pakar dalam pengujian percayakan prestasi berasaskan kebolehan. Kebolehan pula sepatutnya ditunjukkan oleh seseorang individu semasa pengujian dijalankan.

***Strategi Pentaksiran Berasas Prestasi.*** Terdapat banyak strategi yang digunakan untuk mentadbir pentaksiran berasas prestasi seperti senarai semak dan pemerhatian. Pentaksiran ini telah digunakan banyak dekad sebelum ini. Walau bagaimanapun, dengan tujuan yang khusus pentaksiran ini memainkan peranan penting dalam sistem pentaksiran. Strategi pentaksiran yang sesuai digunakan kepada kanak-kanak adalah berbentuk temubual, borang, tugas, permainan dan hasil kerja.

Temubual yang digunakan bukanlah untuk tujuan analisa data dalam pendekatan kualitatif. Tetapi guru menggunakan temubual untuk mengetahui apa yang kanak-kanak faham mengenai sesuatu konsep. Temubual adalah sesuai untuk kanak-kanak yang baru mengembangkan kemahiran literasi dan masih belum boleh



mempamerkannya melalui aktiviti pensel dan kertas. Strategi ini selari dengan teknik temubual yang digunakan oleh Piaget untuk memahami pemikiran kanak-kanak. Melalui pertanyaan dan menyoal lebih soalan berdasarkan maklumbalas atau respon kanak-kanak, Piaget menentukan bukan hanya apa yang kanak-kanak faham, tetapi proses berfikir turut digunakan untuk memberi maklum balas kepada soalan (Seefeldt, 2005).

Temubual terdiri daripada tridak berstruktur, berstruktur atau berbentuk diagnostik. Menurut Wortham (2012), temubual tidak berstruktur boleh dilakukan semasa kanak-kanak bermain, melakukan kerja di pusat-pusat perkembangan, atau sedang berhubung dengan aktiviti di kelas. Pada ketika ini, memberi peluang kepada guru untuk menemubual kanak-kanak dengan soalan yang mencungkil, pengalaman, perasaan dan keupayaannya. Temubual berstruktur pula dirancang untuk mengetahui maklumat yang lebih spesifik agar lebih memahami kanak-kanak. Contoh yang dibawa oleh Engel (1990), selepas sesi bercerita, guru bertanyakan soalan soalan berbentuk penyelesaian masalah untuk mengetahui apa yang ada dalam fikiran kanak-kanak mengenai cerita yang didengarinya. Temubual berbentuk diagnostik pula menyediakan aspek tambahan untuk memenuhi keperluan khusus kanak-kanak dan memahami soalan yang ditanya lebih berfokus dan menentukan keperluan pengajaran. Jika guru telah menyedari kanak-kanak menjadi keliru atau melakukan kesilapan, melalui temubual diagnostik akan mendedahkan pengalaman kanak-kanak melalui respon bertutur.

Borang kontrak mempunyai lebih tujuan iaitu perancangan antara guru dan mencatat untuk hari biasa, sepanjang hari atau beberapa hari. Kanak-kanak TASKA memerlukan gambar atau persembahan visual untuk melengkapkannya. Borang kontrak boleh digunakan sebagai garis panduan oleh guru atau murid terhadap perbincangan kerana ia merupakan satu rekod sistem yang bagus. Selain itu, tugasan

yang khusus memerlukan kefahaman kanak-kanak, lebih baik daripada temubual. Perkara penting di sini adalah guru membuat tugas khusus dan tertentu untuk tujuan penrtaksiran. Perbincangan dan pertanyaan adalah sebahagian daripada proses ini, tetapi kebolehan kanak-kanak menyiapkan tugas adalah menumpukan kepada sesuatu proses (Hills, 1992).

Permainan boleh digunakan untuk memahami kemajuan kanak-kanak melalui kemahiran dan konsep. Walaupun ramai kanak-kanak yang bermain dalam pada masa yang sama, guru boleh menggunakan pemerhatian untuk mentaksir kebolehan dan pemikiran kanak-kanak. Kamii & Rosenblum (1990) mencadangkan guru menggunakan permainan untuk pemerhatian yang sistematik untuk keseluruhan kelas. Permainan boleh digunakan untuk memperkenalkan konsep atau kemahiran di dalam bidang yang berbeza. Sudah berdekad lamanya, melalui permainan kemahiran membaca dikembangkan.

Strategi terakhir adalah menggunakan hasil kerja. Hasil kerja adalah contoh semua jenis kerja dan aktiviti kanak-kanak yang boleh menggambarkan kemajuan perkembangan kanak-kanak. Hasil kerja yang berkaitan seperti hasil capan, asas penulisan, dan konsep bertema seperti yang dinyatakan oleh Ratcliff (2001, 2002). Manakala menurut Grace & Shores (1991) mencadangkan menggunakan media yang lain untuk merekod seperti rakaman video, rakaman gambar atau rakaman audio. Kamera juga dilihat mempunyai fungsi istimewa untuk merekod hasil kerja ini.

***Pengelasan Dan Susun Atur Pentaksiran Berasas Prestasi.*** Pentaksiran berasas prestasi dikenali sebagai pengukuran tidak formal. Walau bagaimanapun, ia boleh dikategorikan berstruktur atau tidak berstruktur dan secara langsung atau tidak langsung. Susun atur ini dilihat sama dengan temubual berstruktur dan tidak berstruktur seperti yang telah dibincangkan sebelum ini dalam strategi pentaksiran prestasi. Kategori pentaksiran ini merujuk kepada jenis aktiviti yang digunakan dalam

instrumen tersebut. Menurut Lee (1992), pentaksiran prestasi tidak berstruktur adalah sebahagian daripada aktiviti pembelajaran di kelas, seperti hasil kerja menulis, projek, senarai semak atau tugas yang diberi guru. Pentaksiran prestasi berstruktur pula adalah yang ditentukan atau direkabentuk termasuk soalan atau tugas yang memerlukan penyelesaian masalah, sintesis dan analisis. Soalan berbentuk terbuka, dan semua murid ditadbir dan melalui proses menjawab soalan yang sama. Pentaksiran berstruktur bukan sahaja telah dirancang, tetapi juga perlu mencapai selari dengan standard untuk kebolehpercayaan dan kesahan, seperti yang diperlukan dan dipenuhi untuk instrumen pentaksiran yang standard. Sebagai contoh, instrumen pentaksiran mempunyai kriteria skor yang tertentu yang benar-benar mengenal pasti tingkah laku yang diukur.

Susun atur pentaksiran berasas prestasi ditentukan berdasarkan objektif dan menentukan jenis tingkah laku yang hendak diukur. Dalam kajian ini, instrumen yang digunakan memerlukan kanak-kanak menunjukkan prestasi dalam pelbagai aktiviti kerana aktiviti yang melibatkan domain fizikal, sosioemosi dan rohani, kanak-kanak diminta melakukan aktiviti yang tiada bukti bertulis, melainkan rekod pemerhatian. Manakala domain kognitif, bahasa dan kreativiti melibatkan aktiviti motor halus dan motor kasar serta pergerakan. Aktiviti menggunting, membuat capan menghasilkan eviden bertulis yang. Maka susun atur pentaksiran ini berkaitan dengan proses pemerhatian dan dokumentasi untuk menghasilkan pentaksiran yang lebih menyeluruh.

Peranan pemerhatian dilakukan untuk mengenali kanak-kanak dan apa yang boleh dilakukannya. Menurut Bergan & Feld (1993), hanya mengetahui apa yang kanak-kanak boleh lakukan tidak mencukupi untuk melaporkan tahap kanak-kanak. Analisis juga perlu dilakukan, maka, pemerhatian yang dilakukan, memainkan peranan penting kepada proses yang seterusnya. Selain itu, kanak-kanak hendaklah

diperhatikan pada waktu tertentu dan menggunakan pelbagai bahan dan alat maujud yang berbeza untuk menentukan maklumat baru telah berkembang.

Peranan penting pemerhatian adalah kepada amalan refleksi di mana guru menggunakan instrumen untuk membuat perancangan kepada kanak-kanak dengan berpandukan hasil pentaksiran. Guru menggunakan masa untuk membuat refleksi terhadap maklumat kanak-kanak yang telah dikumpul. Berdasarkan Hills (1992), langkah yang boleh diikuti oleh guru untuk membuat refleksi daripada data yang diperolehi daripada pemerhatian seperti berikut: (1) menentukan tujuan dan fokus, (2) memerhati dan merekod, (3) menyusun apa yang telah direkod, iaitu secara individu atau berkumpulan dan (4) membuat refleksi apa yang telah direkod dan memfokuskan kembali kepada aktiviti pengajaran dan pembelajaran.

Pemerhatian adalah asas kepada pentaksiran berasas prestasi (Wortham, 2012). Digunakan bersama temubual apabila guru memerhatikan respon dan tingkah laku kanak-kanak. Pemerhatian membolehkan guru memahami fikiran dan pengetahuan kanak-kanak semasa bermain dalam aktiviti yang ditaksir dan diukur. Ia diintegrasikan dengan tugas secara langsung dimana guru memerhatikan kanak-kanak menyiapkan tugas tersebut. Senarai semak dan skor guna skala rating adalah sebahagian daripada proses pemerhatian untuk memahami perkembangan kanak-kanak (Baldwin et al., 2009).

Dalam konteks pentaksiran berasaskan prestasi, dokumentasi membawa peranan yang luas, khususnya apabila dihubungkan dengan program pendidikan awal kanak-kanak dan berpusatkan kanak-kanak atau pengalaman pembelajaran di kelas (Wurm, 2005). Kebiasannya dokumentasi berkait rapat dengan kurikulum pendidikan awal kanak-kanak. Seperti yang dihuraikan oleh Helm & Beneke (2003), Sekolah Reggio Emilia menggunakan Pendekatan Projek dan dokumentasi dalam kurikulum ini adalah proses mendokumenkan kemajuan aktiviti-aktiviti yang telah

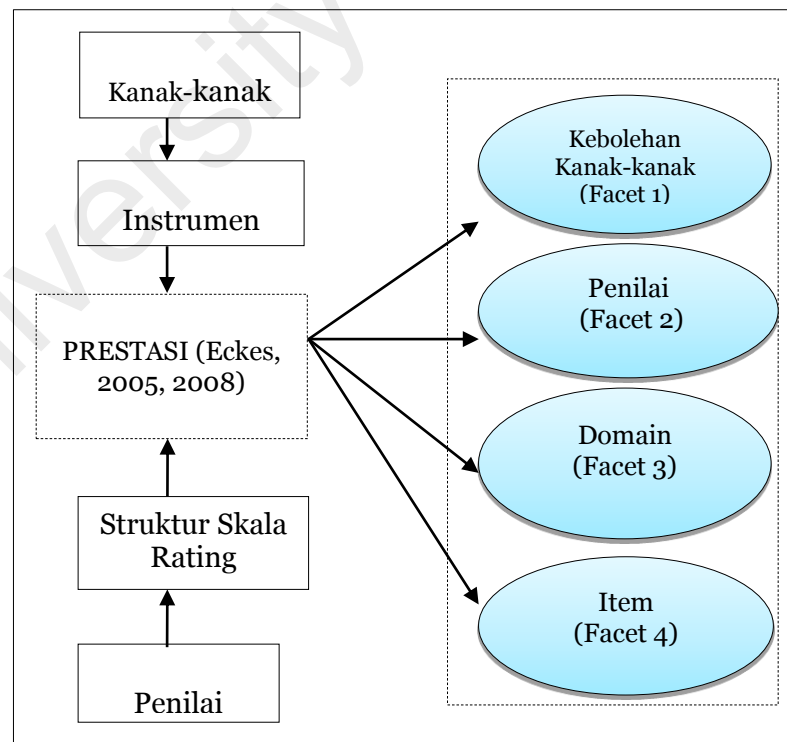
dilaksanakan dan langkah terbaik untuk memahami minat kanak-kanak, pemikiran kanak-kanak dan masalah yang telah diselesaikan dalam aktiviti mereka. Melalui medium pemerhatian, proses dokumentasi dilakukan setelah maklum balas diterima dan ia dilakukan dalam bentuk grafik, atau gambar digital dan hasil kerja kanak-kanak. Ia dipamerkan dan dikongsikan bersama ibu bapa, guru lain selepas habis suatu pentaksiran. Dokumentasi memainkan peranan penting dalam pentaksiran kanak-kanak kerana kebolehan kanak-kanak dilaporkan secara sistematik dan merupakan eviden atau bukti terhadap perkembangan kanak-kanak.

**Pembentukan Konsep Dan Komponen (Pembolehubah) Pentaksiran Berasas Prestasi.** Konsep pentaksiran prestasi sangat luas dan juga mempunyai kanak-kanak dan domain yang berbeza serta mempunyai definisi yang abstrak. Pengkaji akan mengkhususkan konsep-konsep yang abstrak ni supaya dapat mewujudkan item-item yang mentaksir tingkah laku kanak-kanak. Pemberian makna kepada komponen atau pembolehubah kajian serta memberi tafsiran satiap satunya berasaskan teori-teori kognitif dan sosiobudaya serta kajian literature yang dirujuk.

Menurut Eisner (1999), pentaksiran berasas prestasi membawa perkembangan yang sangat penting dalam penilaian. Beliau menghuraikan pentaksiran berbentuk prestasi ini telah mengubah dunia pendidikan kerana ia benar-benar mengenal pasti dan mengukur kebolehan sebenar murid. Pentaksiran ini lebih sesuai dan berkebolehan untuk mentaksir kebolehan kanak-kanak membina pemikiran yang kompleks, bagaimana untuk mengenal pasti masalah, merancang langkah tertentu dan berfokus kepada hasil pembelajaran. Eisner melihat pentaksiran ini adalah pendekatan yang lebih baik daripada ujian pensel dan kertas kerana mentaksir sejauhmana kebolehan kanak-kanak dalam dunia sebenar.

Apabila sejarah pentaksiran berasas prestasi dihuraikan dengan lengkap didalam artikel '*A Short History of Prestasice Assessment* oleh Madaus & O'Dwyer (1999) di dalam Turner, J. (2003), beliau menjelaskan bahawa pentaksiran berasas prestasi adalah pendekatan baru dan autentik untuk menilai murid dalam kemahiran tertentu. Walau bagaimanapun ia telah lama digunakan dalam konteks pendidikan, terutamanya di dalam bidang tertentu atau prestasi yang hendak ditaksir adalah merujuk kepada objektif kurikulum.

Dalam kajian ini, pengkaji menetapkan objektif khusus kajian untuk mentadbir perntaksiran kepada kanak-kanak untuk tindak balas semua komponen dan berupaya mengenal pasti kebolehan kanak-kanak, pengaruh dan bias penilai, kesukaran domain dan kesukaran item yang dianalisis menggunakan Model Pengukuran Moden Rasch-Facet, yang juga merupakan pemboleh ubah kajian yang berwarna biru di dalam Rajah 2.5 di bawah:



Rajah 2.5. Komponen yang ditaksir

Menurut McNamara (1996), komponen yang terlibat dengan pentaksiran berasas prestasi adalah (berdasarkan hala anak panah), kanak-kanak yang menunjukkan prestasi, berdasarkan instrumen yang digunakan, sebelum ditaksir berdasarkan objektif yang ditetapkan. Penilai pula berhubung dengan bentuk penskoran untuk mentaksir prestasi. Interaksi yang wujud di sini adalah antara penilai dengan kanak-kanak semasa pentaksiran berlangsung.

**Kebolehan Kanak-kanak.** Kebolehan kanak-kanak bermaksud sesuatu tahap yang dicapai dalam melakukan sesuatu aktiviti atau perkara. Kebolehan dan perkembangan kanak-kanak mempengaruhi kepelbagaian tahap prestasi kanak-kanak ke atas tugas (item). Kanak-kanak di dalam kumpulan umur yang sama dilihat mempunyai Zon Perkembangan Proksimal yang berbeza iaitu; Tugas dilihat mudah bagi sesetengah kanak-kanak dan sedikit mencabar atau sukar atau mustahil dilakukan bagi sesetengah kanak-kanak yang lain (Ormrod, 2011). Kebolehan kanak-kanak mengikut umur 2 dan 3 serta umur 3 dan 4 juga dilihat menggunakan perbezaan *mean measure* dan *boxplot*.

Tambahan juga, kanak-kanak ini datang daripada latar belakang yang berbeza dari segi pengetahuan, kemahiran dan cara berfikir. Bergantung kepada pengetahuan tertentu dan tahap kebolehan, kanak-kanak yang berada dalam kumpulan umur yang berbeza memerlukan *scaffolding* atau sokongan untuk menjayakan sesuatu tugas. (Lodewyk & Winne, 2005).

**Pengaruh Penilai.** Penilai bermaksud individu yang memberikan skor secara efisien dan berkait rapat untuk menterjemah, mentaksir maklum balas oleh kanak-kanak. Pengaruh penilai bermaksud kepelbagaian ciri-ciri yang dimiliki penilai itu sendiri yang menentukan tahap *severity* atau *leniency* semasa membuat pentaksiran (Myford & Wolfe, 2003). Kepelbagaian ciri ini juga boleh mengganggu kesahan dan keadilan pentaksiran berasas prestasi (Lane & Stone, 2006).

Bias penilai pula antara istilah yang berkait rapat dengan pengaruh penilai (Johnson, Penny & Gordon, 2009). Pendekatan yang digunakan untuk mengatasi masalah ini adalah melalui 3 cara iaitu, latihan penilai, mentadbir skor kepada item prestasi yang sama oleh dua orang penilai atau lebih (Myford & Wolfe, 2002) dan menstabilkan kebolehpercayaan *inter-rater* (Kondo-Brown, 2002). Pengukuhan antara penilai dilakukan selepas kajian sebenar dilaksanakan antara dua lokasi menggunakan teori pengukuran moden Rasch-Facet berdasarkan indeks pengukuhan yang dianalisis (Linacre, 2008).

Terdapat banyak kajian menggunakan pendekatan penilaian berasaskan prestasi yang mengambil kira faktor ketegasan penilai (Engelhard, 1994, 1996, 1997; Engelhard & Myford, 2003; Lunz, Wright, & Linacre, 1990; Lunz & O'Neil, 1996; MacMilian, 2000; Myford & Wolfe, 2002). Prosedur ini membenarkan penyelarasan *raw scores* dengan mengenal pasti penilai sebagai *severe* atau *lenient* dan seterusnya melaraskan skor prestasi. Berdasarkan kajian Lunz dan Schumacker (1997), perbandingan metodologi digunakan untuk mendapatkan ketepatan yang lebih tinggi melalui rating yang diberi oleh penilai. Metodologi yang dibandingkan adalah (a) *traditional summary statistic*, (b) *inter-judge correlations*, (c) *Generalizability theory*, dan (d) PRPF. PRPF digunakan untuk mengesan *rater severity* dalam menentukan keputusan lulus-gagal. Metodologi ini membolehkan penilai mengambil kira kesan interaksi bias semasa proses penilaian dan kebolehan sebenar sampel dikenal pasti dengan lebih baik.

Melalui kajian ini, metodologi model pengukuran Rasch pelbagai-faset dan Theory G (*Generalizability theory*) akan digunakan. Metodologi ini dapat dimanfaatkan oleh penyelidik-penyelidik dalam bidang pendidikan awal kanak-kanak dalam merancang, melaksanakan dan mentaksir perkembangan kanak-kanak yang mempertimbang untuk menggunakan penilai dan untuk mengetahui jumlah penilai



yang sesuai untuk menjalankan pentaksiran. Hasil kajian ini dapat menunjukkan hasil analisis kerja dalam portfolio kanak-kanak dinilai secara kuantitatif, dipercayai, sah dan objektif.

**Faset Domain.** Mengenal pasti item-item di dalam domain ini mengukur domain yang sama iaitu merujuk kepada ciri unidimensionaliti. Disamping mendapatkan nilai kebolehpercayaan bagi keenam-enam domain.

**Kesukaran Item.** Kesukaran item diterjemah menggunakan indeks infit dan outfit yang didefinisikan sebagai prestasi individu kanak-kanak yang memperolehi nilai positif menunjukkan item tersebut sukar kepadanya. Jika nilai negatif, menunjukkan domain adalah lebih mudah (Linacre, 2002).

**Menghubungkan Standard dan Pentaksiran Berasaskan Prestasi.** Standard untuk pencapaian pembelajaran kanak-kanak adalah dihubungkan dengan tetapan ujian standard dalam dasar pendidikan sesebuah negara. Bagaimana standard ini memberi kesan kepada pembelajaran autentik dan prestasi? Pentaksiran berasas prestasi adalah penting semasa memahami perkembangan dalam pendidikan awal kanak-kanak (Wortham, 2012). Untuk memahami standard dan pentaksiran berasas prestasi, guru atau sesiapa yang bertanggungjawab menjalankan pentaksiran, mesti memahami bagaimana standard diintegrasikan ke dalam kurikulum dan bagaimana keduanya bergabung dari segi pelaksanaannya. Dengan mengenal pasti standard, guru berupaya melengkapkan dan menjadikan pentaksiran amalan terbaik untuk menguji keberkesanan standard melalui prestasi yang ditaksir.

Berdasarkan kajian ini, pentaksiran item prestasi kanak-kanak adalah selari dengan standard kurikulum kebangsaan iaitu sesuai dengan standars pentaksiran kanak-kanak dengan tujuan, hasil kajian ini dapat digunakan sebagai panduan di Taska-taska di seluruh Malaysia dan meningkatkan lagi amalan refleksi guru melalui strategi pentaksiran sekaligus mendokumentasikan hasil pentaksiran kebolehan

sebenarnya kanak-kanak. Oleh yang demikian, hubungan dasar kementerian dan pihak pelaksana iaitu di TASKA akan menjadi baik hubungan secara dua hala demi pembangunan kanak-kanak Malaysia yang berkesan.

**Menghubungkan Standard Kepada Pembelajaran Autentik.** Pembelajaran Autentik adalah berkait dengan aktiviti yang mempamerkan prestasi dalam konteks sebenar. Perkara pertama dalam menghubungkan standard dengan pentaksiran berasaskan prestasi adalah untuk menghubungkan standard dengan kurikulum. Ia bertujuan untuk membina hubungan antara standard dan amalan terbaik untuk kanak-kanak seawal usianya demi kualiti program pendidikan awal kanak-kanak yang melalui pelbagai proses. Tujuan ini turut disokong oleh kajian oleh Baldwin (2009), Drew, Christie, Johnson, Meckley, & Nell (2008) yang menyatakan proses yang berlaku seperti pembelajaran berpusat kanak-kanak, pembelajaran aktif dengan persekitaran yang banyak memberi peluang kepada kanak-kanak untuk melakukan pelbagai aktiviti, pembelajaran dan permainan dalaman serta luaran.

Menurut Kajian Drew (2008), standard boleh dihubungkan dengan kurikulum antaranya melalui aktiviti bermain seperti yang telah dilakukan oleh sekumpulan pengkaji dalam bidang pendidikan, iaitu mereka berjaya memadankan Standard Awal Pembelajaran Arizona melalui kajian terhadap *constructive play*. Standard kebangsaan boleh dipadankan dengan topik yang dirancang (Jacobs & Crowley, 2010).

**Menghubungkan Standard Kepada Pentaksiran Prestasi.** Pentaksiran boleh dihubungkan kepada standard dengan menggabungkan kesemua strategi pentaksiran berasaskan prestasi iaitu menggunakan temubual, pentaksiran secara langsung, borang prestasi, permainan dan hasil kerja. Melalui eviden atau bukti yang diperolehi seperti tingkah laku, interaksi, dan hasil kerja kanak-kanak, maka secara langsung, prestasi kanak-kanak dihubungkan dengan standard kebangsaan (Baldwin, 2009). Pembelajaran adalah satu proses, kanak-kanak akan melalui pelbagai

pengalaman aktiviti pembelajaran sebelum mencapai tahap standard yang ditetapkan. Kepelbagaian perkembangan kanak-kanak memberi kesan kepada pencapaian mereka semasa melakukan aktiviti. Kepelbagaian ini akan memberi pelbagai hasil pemerhatian yang memberi respon kepada item dan dapat diselarikan dengan standard.

### **Rumusan**

Pentaksiran berasas prestasi sangat relevan dalam bidang kanak-kanak apabila mereka diberikan peluang untuk menunjukkan kebolehan, maka prestasi yang diperhatikan adalah lebih semulajadi dan autentik atau sebenar. Pentaksiran ini dilakukan menggunakan pelbagai strategi dan strategi ini diselarikan dengan objektif kajian agar prestasi yang ditaksir mampu menunjukkan perbezaan kebolehan antara kanak-kanak, menunjukkan item atau domain yang lebih sukar dan pengaruh oleh penilai yang membuat pemerhatian. Pentaksiran ini selari dengan standard kurikulum kanak-kanak di Malaysia dan mencakupi kepelbagaian perkembangan kanak-kanak melalui responden yang dipilih adalah pelbagai

## **Bab 3 Metodologi Kajian**

### **Pengenalan**

Bab ini menerangkan metodologi kajian bagi mencapai objektif kajian, iaitu menghasilkan instrumen yang sah dan boleh dipercayai untuk mengukur prestasi kanak-kanak berumur dua hingga empat tahun berpandukan rubrik standard prestasi berdasarkan penilaian guru. Bab ini menghuraikan rekabentuk kajian, responden kajian, prosedur pengumpulan data, prosedur penganalisan data dan rumusan. Rekabentuk kajian dibahagikan kepada tiga bahagian iaitu peringkat rekabentuk, peringkat pembinaan dan peringkat pengesahan. Setiap peringkat dibincangkan secara terperinci. Seterusnya, penyelidik membincangkan responden kajian yang terlibat bagi kajian ini diikuti dengan prosedur pengumpulan data. Prosedur penganalisan data pula dibincangkan menurut persoalan kajian.

### **Reka Bentuk Kajian**

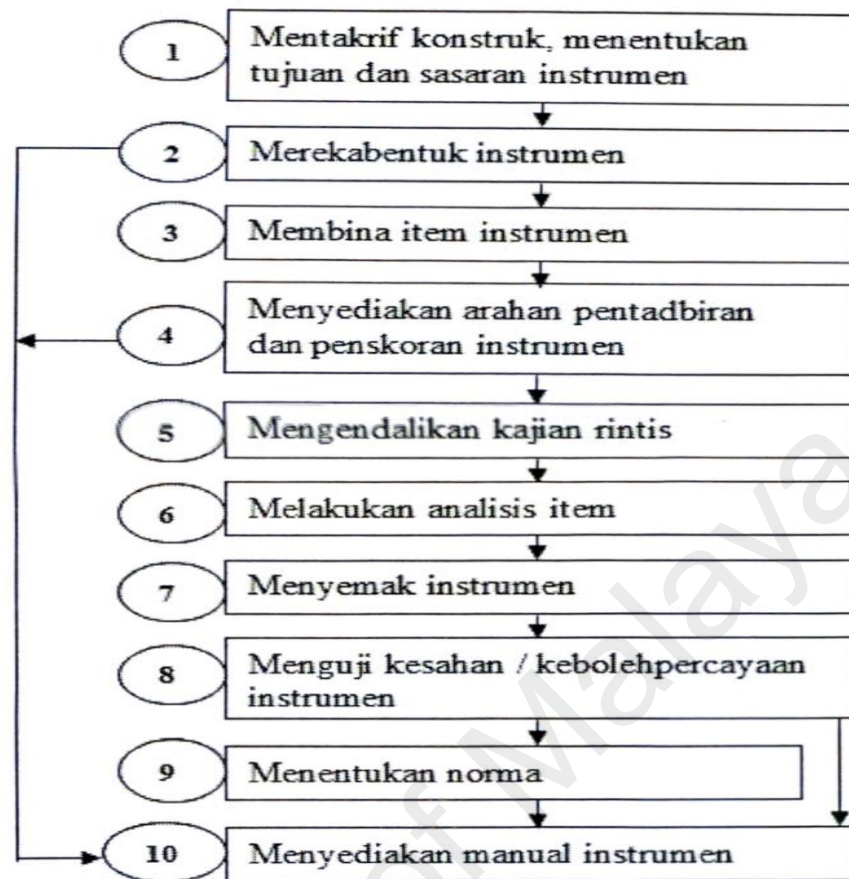
Reka bentuk kajian melibatkan pendekatan kuantitatif yang disepadukan dalam tiga peringkat yang tersusun mengikut satu prosedur yang berurutan daripada awal hingga akhir. Ketiga-tiga peringkat tersebut adalah Peringkat Mereka bentuk, Peringkat Membina Instrumen dan Peringkat Mentaksir. Proses mereka bentuk dilakukan selari dengan prosedur yang telah ditetapkan dalam Modul Pentaksiran Perkembangan Kanak-kanak yang dibina oleh Bahagian Pembangunan Kurikulum, Kementerian Pelajaran Malaysia pada tahun 2010. Prosedur ini diambil kira kerana modul pentaksiran ini adalah satu-satunya panduan yang dihasilkan mengikut budaya kanak-kanak Malaysia dan ia merupakan sistem yang telah dilaksanakan untuk mentaksir kanak-kanak di Malaysia.

Instrumen Standard Awal Pembelajaran Perkembangan (ELDS) yang dibina berasaskan budaya Malaysia telah diimplementasi kali pertama pada tahun 2010 terhadap kanak-kanak berumur dua hingga empat tahun di seluruh negeri di Malaysia. Pendekatan pentaksiran pembelajaran dan perkembangan kanak-kanak dalam kajian ini telah direkabentuk semula oleh penyelidik pada tahun 2013 untuk memasukkan prosedur penskoran secara skala likert (lima rating) yang berbeza dengan kaedah penskoran instrumen pentaksiran terdahulu (dikotomos; ya atau tidak). Pengumpulan data pada tahun 2013 bertujuan untuk kajian rintis manakala pengumpulan data pada tahun 2014 adalah untuk kajian sebenar dan disertasi. Berdasarkan Model Miller, Lovler & McIntire (2013), diperincikan langkah-langkah yang terlibat bagi setiap peringkat bermula peringkat merancang, peringkat rekabentuk dan pembinaan serta peringkat pengesahan.

### **Model Pembinaan Instrumen Miller (2013)**

Selepas merancang, Model Miller, Lovler & McIntire (2013) digunakan dengan gabungan langkah pembentukan instrumen oleh Turner (2001) dan Wortham (2012). 13 langkah pembentukan instrumen ini digunakan bagi kajian ini. Penyelidik membahagikan langkah-langkah pembinaan instrumen ini kepada dua peringkat, iaitu peringkat reka bentuk dan pembinaan agar kerangka pembinaan instrumen ini lebih sistematik. Peringkat terakhir adalah peringkat pengesahan

Fasa 1	Peringkat Rekabentuk Peringkat Pembinaan	McNamara (1996) Turner (1999) Wortham (2012)
Fasa 2	Peringkat pengesahan	Model PRPF dan EduG



Rajah 3.1. Model Pembinaan Instrumen (Miller, Lovler & McIntire, 2013).

**Peringkat Rekabentuk.** Sebelum pentaksiran dilakukan, penyelidikan membuat perancangan dengan menentukan komponen-komponen penting yang ingin dikaji seperti yang dicadangkan oleh McNamara (1996) seperti berikut.

**Kumpulan Sasaran.** Instrumen yang diadaptasi disesuaikan untuk kanak-kanak berumur 2 hingga 4 tahun. Untuk tujuan kajian ini, proses pengujian akan ditadbir kepada kanak-kanak di Taman Asuhan Kanak-kanak (TASKA) swasta dengan mengambil kira komposisi jantina dan lokasi TASKA.

**Tujuan Mengadaptasi Instrumen.** Instrumen yang diadaptasi adalah menggunakan instrumen yang asalnya dihasilkan untuk kumpulan umur awal kanak-kanak di Malaysia yang terbahagi kepada dua instrumen yang ditadbir kepada kanak-kanak 2 hingga 4 tahun. Instrumen ini diadaptasi dengan tujuan untuk mentaksir dan

menentukan kebolehan kanak-kanak berdasarkan prestasinya ke atas setiap item dalam domain serta menentukan item dan domain mana yang sukar bagi kanak-kanak kerana instrumen asal mengambil kira secara keseluruhan apa yang boleh dilakukan oleh kanak-kanak. Dengan mengambil kira faktor umur, jantina dan lokasi, pentaksiran menggunakan instrumen yang telah diadaptasi dapat menentukan kedudukan dan membezakan kebolehan kanak-kanak 2 hingga 4 tahun dengan tepat.

***Menentukan Skala.*** Penyelidik menggunakan skor secara skala rating (skor 1,2,3,4,dan 5) yang boleh membezakan kebolehan kanak-kanak kerana kebolehan kanak-kanak melakukan tugas item) adalah berbeza samada dengan bantuan, tanpa bantuan atau berkebolehan melakukannya dengan sendiri. Skor yang diberikan mewakili peningkatan terhadap kebolehan kanak-kanak.

***Menggunakan Penilai.*** Penyelidik menggunakan pemerhatian oleh penilai kerana sampel kajian adalah kanak-kanak yang tidak boleh menilai diri mereka sendiri. Strategi pemarkahan secara keseluruhan bergantung kepada pemerhatian subjektif oleh penilai iaitu individu yang dilantik, diberi latihan khas dan melaksanakan proses pemerhatian terhadap sampel dan mencatat skor. Dua orang penilai akan memerhati sampel yang sama dan memberi skor dalam kajian rintis.

***Prestasi Tugas (item).*** Setiap domain dalam instrumen ini adalah berbentuk item prestasi di mana kanak-kanak akan menunjukkan kebolehannya melakukan tugas atau menyelesaikan masalah samada tanpa bantuan atau dengan bantuan.

## **Peringkat Pembinaan Instrumen**

*Mengenal pasti Standard Pentaksiran.* Standard Pentaksiran Perkembangan Murid KPM dirujuk untuk melakukan penambahan ke dalam instrumen. Perkara yang ditambah adalah Fokus, Standard Kandungan dan Standard Pembelajaran manakala item prestasi dan aktiviti adalah daripada instrumen asal. Penambahan ini boleh dirujuk dalam Instrumen IPHKT di lampiran.

Fokus adalah bidang pembelajaran yang ingin dikembangkan dalam diri kanak-kanak. Fokus dinyatakan melalui Standard Kandungan. Standard Kandungan merupakan pernyataan spesifik tentang perkara yang kanak-kanak patut tahu dan boleh lakukan. Setiap Standard Kandungan diperincikan lagi kepada Standard pembelajaran. Standard Pembelajaran merupakan satu penetapan kriteria atau indikator pencapaian yang boleh diukur, untuk memastikan penguasaan setiap Standard Kandungan.

*Mereka bentuk Instrumen Selepas Diadaptasi.* Instrumen Standard Awal Pembelajaran Perkembanagan kanak-kanak direka bentuk semula untuk mematuhi konsep pentaksiran Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan, dan bermula tahun 2010 telah digunapakai di TASKA swasta berkaitan perkara tersebut iaitu:

Perkara 21- Konsep kurikulum kebangsaan (i, iii dan vii) iaitu;

i) pendidikan secara holistik atau *Whole brain learning*;

iii) pembelajaran adalah menyeronokkan (*learning is fun*), bertujuan membina kemahiran sosial, kemahiran awal matematik dan awal literasi, merangsang minda dan meningkatkan rasa ingin tahu (*curiosity*) yang tinggi, di samping membina sosio-emosi yang ceria dan tenang;



vii) Penilaian kanak-kanak adalah berasaskan pemerhatian klinikal menggunakan kaedah 'pedagogi permainan' (play pedagogy) yang direkodkan dengan jelas.

Perkara 23- Indikator yang boleh diambil kira bagi menentukan kualiti Kurikulum Kebangsaan (vii) iaitu;

vii) penaksiran perkembangan kanak-kanak dibuat melalui pelbagai teknik yang menggambarkan prestasi kanak-kanak yang sebenar, yang bermaksud penilaian dibuat berasaskan sumber-sumber yang terbukti kesahihannya dan juga dari bukti-bukti yang diperolehi daripada persekitaran yang sebenar dan daripada pelbagai situasi yang mencerminkan prestasi sebenar kanak-kanak.

***Mentaksir Prestasi.*** Semasa membina instrumen ini, objektif kajian ditentukan untuk menguji instrumen prestasi yang telah diadaptasi dengan mentaksir empat perkara berikut (Eckes, 2005; 2008):

***Kebolehan kanak-kanak.*** Antara kanak-kanak dalam kumpulan umur 2 dan 3 tahun, terdapat kanak-kanak yang berkebolehan tinggi dan rendah melakukan tugas (item). Kedudukan ini yang perlu ditentukan.

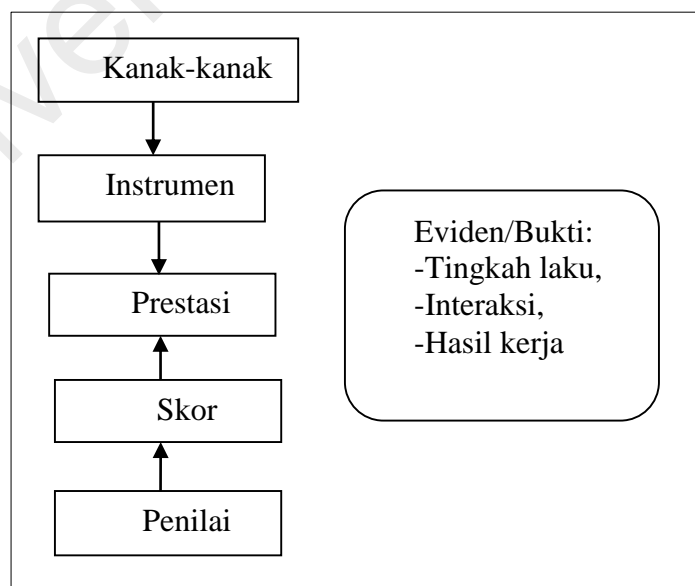
***Pengaruh Penilai.*** Penilai adalah individu yang memberi skor. Maka, penilai yang memiliki ketegasan dan kelembutan (*severity and leniency*) perlu ditentukan. penilai turut menilai prestasi kanak-kanak terhadap domain-domain instrumen, samada penilai ini mempunyai bias atau tidak, maka indeks bias antara penilai dengan kumpulan umur perlu ditentukan.

***Faset Domain.*** Domain pembelajaran sangat penting kepada kanak-kanak. Maka ciri unidimensionaliti bagi setiap domain perlu dicari agar ia menepati andaian Rasch yang mengatakan bahawa item-item dalam domain ini mengukur satu

keupayaan yang tunggal (Wright & Masters, 1982). Maka, domain-domain yang terlibat adalah Domain Kognitif, Domain Bahasa, Domain Kreativiti, Domain Fizikal, Domain Sosioemosi atau Domain Rohani.

*Kesukaran Item.* Pencapaian kanak-kanak ditentukan melalui prestasi yang ditunjukkan semasa melakukan aktiviti atau tugas (item). Terdapat kanak-kanak yang boleh melakukan dengan mudah, seperti kanak-kanak berumur 2 tahun boleh melakukan sama seperti kanak-kanak 3 tahun lakukan. Terdapat item yang sukar. Maka, item yang sukar bagi kanak-kanak perlu ditentukan.

Kajian ini menekankan pentaksiran prestasi kepada semua domain instrumen (holistik) dalam suasana pembelajaran yang menyeronokkan (*learning is fun*) yang melibatkan aktiviti pembelajaran bersama guru di taska. Mentaksir adalah untuk mengumpul maklumat. Kaedah pemerhatian digunakan untuk mengumpul maklumat atau evidens. Pemerhatian bermaksud pengamatan kepada perlakuan kanak-kanak. Kajian ini menggunakan peranan penilai untuk membuat pemerhatian terhadap kanak-kanak seperti Rajah 3.2 di bawah menunjukkan proses pentaksiran instrumen prestasi menggunakan penilai.



Rajah 3.2 Proses pentaksiran menggunakan penilai oleh McNamara (1996)

Perkara yang diperhatikan ialah eviden atau bukti. Bukti prestasi yang boleh dipamerkan dalam bentuk tingkah laku, interaksi, hasil kerja. Bukti yang mahu dilihat oleh penilai ialah yang menunjukkan perkembangan dan pencapaian kanak-kanak daripada segi tingkah laku, interaksi dan hasil kerja kanak-kanak. Penilai akan diberi latihan agar tiada beza penskoran yang ketara antara mereka dan untuk meningkatkan konsistensi antara keduanya semasa pentaksiran.

***Penghakiman Panel Pakar dan Semakan Instrumen.*** Selepas instrumen telah siap direka bentuk, pengkaji menghantar ke pakar untuk mendapatkan maklum balas terhadap kesahan kandungan instrumen. Instrumen dinilai oleh panel pakar dalam bidang-bidang yang sesuai dengan pendidikan awal kanak dan dari segi bahasa. Instrumen yang digunakan dalam kajian ini, telah dibina mengikut budaya kanak-kanak Malaysia dan dalam Bahasa Malaysia. Maka pengkaji membuat kesahan kandungan sahaja untuk melihat maklum balas daripada pakar. Pengkaji membuat semakan ke atas item-item instrumen yang telah dinilai oleh pakar untuk memastikan instrumen sah digunakan selepas diadaptasi.

***Kajian Rintis.*** Apabila selesai membuat semakan, item-item kemudian ditadbirkan kepada kumpulan yang mempunyai latar belakang yang sama dengan kumpulan sasaran dalam kajian sebenar. Semasa mentadbir instrumen dijalankan, pengkaji melihat bentuk item dan tingkah laku yang dicapai oleh kanak-kanak untuk membaiki kelemahan. Masa yang diambil semasa mentadbir instrumen juga direkodkan untuk tujuan pengubahsuaian. Kajian rintis akan melaporkan nilai kebolehpercayaan.

***Menganalisa Data Rintis dan Data Latihan Guru.*** Tujuan menganalisa data pada peringkat ini adalah untuk melihat nilai kebolehpercayaan instrumen selepas diadaptasi dan melaporkannya kerana menggunakan instrumen yang diadaptasi.

Penyelidik melaporkan hasil analisis menggunakan Model PRPF. Analisis juga melibatkan indeks kebolehan kanak-kanak, indeks ketegasan penilaian, kesukaran item, domain dan bacaan indeks bias penilai.

### **Peringkat Pengesahan**

Kajian dijalankan agar instrumen yang dibina dapat dilihat akan kesesuaiannya untuk mengukur perkembangan kanak-kanak. Menurut Cohen dan Swerdlik (2005) instrumen yang dibina perlu dicuba kepada situasi yang sama atau hampir sama dengan situasi di mana kajian tersebut akan dijalankan seperti arahan yang digunakan, masa yang ditetapkan dan suasana bagaimana instrumen tersebut akan dijalankan.

*Analisis Data.* Dalam menentukan kesahan dalam menggunakan instrumen pembelajaran, perkembangan awal kanak-kanak yang dibina, penyelidik menggunakan model PRPF sebagai model utama, dan Teori Generalizabiliti melalui perisian Edu-G sebagai pelengkap analisis. Penggunaan model ini melibatkan empat faset utama iaitu kanak-kanak, penilai, domain dan item. Penyelidik mengumpul bukti-bukti pengesahan penggunaan instrumen ini dengan menggunakan bukti pengesahan yang disarankan oleh *Standards for Educational and Psychological Testing* oleh *American Educational Research Association, American Psychological Association* dan *National on Measurement in Education* (1999). Bukti-bukti pengesahan ini telah dibincangkan di Bab 2. Seterusnya penyelidik melihat konsistensi penilai melalui kajian terhadap nilai bias.

Proses pengesahan bagi memastikan instrumen ini sah dan boleh dipercayai dijalankan dalam dua pusingan sepertimana yang disarankan oleh Kane (2006b). Pusingan pertama fokus kepada pembinaan instrumen pengukuran yang mempunyai ciri-ciri-psikometrik yang bersesuaian dengan standard perkembangan kanak-kanak

serta melihat sama ada instrumen yang dibina berfungsi sepertimana yang dijangkakan. Pada pusingan kedua pula,

### **Responden Kajian**

Populasi untuk pengujian instrumen terdiri daripada Jumlah Taska Swasta yang berada di daerah Hulu Selangor (rintis) dan Shah Alam di Selangor. Sampel kanak-kanak berumur 2, 3 dan 4 tahun dipilih secara bertujuan untuk memastikan jumlah sampel mencukupi untuk kajian rintis dan sebenar. Untuk kajian rintis, sampel dipilih seramai 120 orang. Instrumen 1 ditadbir kepada 60 orang kanak-kanak 2 tahun dan 3 tahun. Instrumen 2 juga ditadbir kepada 60 kanak-kanak 3 tahun dan 4 tahun. Jumlah keseluruhan adalah seramai 120 orang. Jumlah ini ditentukan untuk mematuhi prosedur mencari nilai kebolehppercayaan instrumen iaitu 30 orang bagi setiap kumpulan yang sama semasa kajian rintis.

Untuk kajian sebenar di Shah Alam, Jumlah keseluruhan sampel kanak-kanak berumur 2 hingga 4 tahun adalah seramai 115 orang. Mengikut model pengukuran Rasch, 30 orang kanak-kanak yang mewakili setiap kumpulan umur adalah bersesuaian dengan nilai logit 1 (*confident* 95%). Pemilihan kedua-dua lokasi ini adalah untuk mendapatkan kepelbagaian sampel dan Shah Alam adalah kawasan yang mempunyai paling banyak taska.

### **Prosedur Pengumpulan Data**

Sebelum kajian rintis dan kajian sebenar dijalankan, penyelidik terlebih dahulu memohon kebenaran pengusaha TASKA Swasta untuk menjalankan penyelidikan di semua TASKA yang terlibat menggunakan Instrumen Standard Awal Pembelajaran dan Perkembangan.

Pada peringkat perancangan, kumpulan sasaran kanak-kanak terdiri kepada dua kumpulan iaitu Kumpulan 1 berumur 2 dan 3 tahun manakala Kumpulan 2 iaitu umur 3 dan 4 tahun di Taman Asuhan Kanak-kanak (TASKA) swasta. Dalam kajian rintis, dua orang penilai juga dilantik sebagai alat kajian yang melakukan pemerhatian sampel kajian di semua TASKA Hulu Selangor yang terlibat.

Pada peringkat pembinaan instrumen, instrumen telah direkabentuk semula dari segi susun atur kandungan instrumen, menggunakan skor skala rating dan menyusun atur domain-domain instrumen serta mendapatkan kesahan kandungan daripada panel pakar. Kajian rintis dijalankan untuk mendapatkan nilai kesahan dan kebolehpercayaan dan melaporkannya. Semasa peringkat pembinaan, objektif kajian diselarikan dengan enam perkara yang ditaksir menggunakan instrumen ini.

Pada peringkat pentaksiran, kaedah pemerhatian digunakan untuk mendapatkan bukti daripada prestasi yang ditunjukkan oleh kanak-kanak. Pemerhatian dilakukan oleh penilai dan menggunakan borang untuk menentukan skala kebolehan kanak-kanak daripada 1 hingga 5. Bagi mengurangkan sumber ralat atau jarak yang ketara antara kedua penilai, latihan kepada penilai telah diberikan untuk meningkatkan keseragaman nilai skor sesama mereka semasa mentaksir.

### **Prosedur Penganalisan Data**

Model PRPF digunakan sebagai kerangka statistik untuk menganalisis data daripada proses pentaksiran pembelajaran dan perkembangan kanak-kanak. Membuat keputusan berdasarkan data mentah yang dikumpul daripada skor prestasi tidak dapat menghuraikan kesukaran item dan suatu domain, ketegasan setiap penilai, dan tahap dimensi skala rating. Bagaimana kebolehan kanak-kanak dapat ditentukan dengan tepat, jika ketegasan penilai terhadap domain yang sukar menentukan kanak-kanak A, dan kelembutan penilai terhadap domain yang mudah menentukan skor kanak-kanak

B? Linacre (1994) menyatakan “Bagaimana pengukuran yang adil dan bermakna boleh diperolehi daripada skor rating ordinal yang belum pasti” Model PRPF telah menetapkan apa ciri-ciri data yang mesti ada untuk menyokong pengukuran yang bermakna. Apabila data tidak fit dengan model, makna sesuatu pengukuran tidak jelas atau meragukan dan memerlukan penyelidikan lagi.

Adalah menjadi keperluan semasa melakukan pemerhatian sebenar, skor persembahan kanak-kanak diberikan oleh penilai, menjadi lebih teratur kepada konstruk atau ciri terpendam (kebolehan kanak-kanak). Semakin tinggi skor, menunjukkan semakin besar jumlah ciri terpendam yang diukur. PRPF berupaya menunjukkan nilai indeks untuk mengesan keadaan di mana keperluan data tidak dipenuhi jika mengikut kaedah kualitatif. Fit statistik dapat menunjukkan dan mengenal pasti data yang berfungsi atau tidak berfungsi seperti yang sepatutnya. Indikator PRPF menyediakan maklumat jelas dalam pengukuran, sama ada domain pembelajaran atau perkembangan terlalu sukar atau mudah, adakah penilai tegas atau sebaliknya, kriteria yang dimiliki penilai dalam keadaan yang konsisten, adakah kanak-kanak diberi skor tinggi kerana penilai bersifat pemurah atau domain itu senang, kesemua soalan ini dapat dijawab semasa menggunakan model ini.

Model pengukuran Rasch menyediakan maklumat bagi pengkaji untuk ‘mengukur konstruk dengan selari daripada pemerhatian kualitatif’. Dengan menggunakan pentaksiran berasas prestasi, seperti hasil kerja, inilah model yang paling tepat diperlukan. Menggunakan penilai untuk membuat pengukuran subjektif semasa pemerhatian dan terhadap hasil kerja kanak-kanak yang memperlihatkan elemen penilaian kualitatif boleh ditukar kepada skor kuantitatif. Jika objektif untuk menentukan dan mengukur kebolehan sebenar kanak-kanak dan potensi dalam dirinya, maka sangat menekankan kepada keperluan melihat interaksi antara penilai dengan

kesukaran domain terhadap kanak-kanak, melihat interaksi ketegasan penilai dengan kanak-kanak, dan skala rating yang digunakan oleh penilai. Perbandingan kuantitatif oleh PRPF tidak bergantung ke atas siapa kanak-kanak yang ditaksir oleh mana-mana penilai ke atas apa domain pembelajaran dan perkembangan, atau dimensi rating yang digunakan tetapi melihat perbandingan secara individu dengan data yang dianalisis setiap satu bagi setiap kanak-kanak, setiap penilai dan setiap item.

Kajian ini melibatkan 4 faset dalam Model Pengukuran Rasch Pelbagai-Faset yang dikenali sebagai (a) kebolehan kanak-kanak, (b) pengaruh penilai, (c) kesukaran domain dan (d) kesukaran item. Hasil dapatan analisis akan digeneralisasi merentasi keseluruhan kanak-kanak. Generalisasi ini memerlukan pembinaan proses pengukuran iaitu mengukur kebolehan kanak-kanak yang sepadan dengan item, domain, penilai dan rating. PRPF memisahkan perkiraan anggaran bagi ukuran kanak-kanak daripada anggaran kesukaran domain dan ia dilanjutkan untuk memasukkan anggaran berasingan ketegasan penilai dan bias penilai. Keadaan ini membenarkan anggaran ukuran kanak-kanak menjadi ujian tersendiri dan tidak bersandar.

Struktur skala rating  $Fk$  tidak dipertimbangkan sebagai pemboleh ubah *faset* kerana skala rating adalah termasuk dalam format atau struktur keseluruhan. Menambah kategori lain akan mengubah hubungan kategori yang sedia ada. Elemen boleh ditambah dengan untuk menentukan ia untuk menjadi satu *faset*. Menempatkan kanak-kanak, item, domain, penilai dan dimensi rating pada satu garisan linear menyediakan satu kerangka rujukan untuk mengenal pasti hubungan setiap *faset* dalam pentaksiran pembelajaran dan perkembangan kanak-kanak. Berdasarkan kerangka tersebut, boleh melihat anggaran kebolehan kanak-kanak daripada tinggi ke rendah, anggaran kesukaran item dan domain iaitu dari yang paling sukar hingga mudah,



anggaran ketegasan penilai iaitu daripada yang paling tegas hingga lembut, dan menganggarkan kesukaran domain iaitu yang sukar hingga mudah.

Pengiraan kuantitatif setiap faset adalah dalam logit yang menempatkan ukuran (kanak-kanak, penilai, domain dan item) di atas satu garis skala linear yang dipanggil *variable map*. Setiap faset ditentu ukur daripada skor pemerhatian prestasi kanak-kanak yang berpusat pada nilai '0'. Model *Four-faset Rasch* dihuraikan sebagai *log-odds* kepada kebolehan kanak-kanak  $B_n$ ,

Example: "Suzie ( $n$ , person) is rated as 3( $k$ , category) by Prof. Chin ( $j$ , judge) on grammar ( $i$ , item) in Essay C ( $m$ , task)

$$\log_e(P_{njimk}/P_{njim(k-1)}) = B_n - D_i - S_j - T_m - F_k$$

di mn:

- $P_{njimk}$  - is probability of being rated in category  $k$
- $P_{njim(k-1)}$  - is the probability of being rated in category  $k-1$
- $B_n$  - ability of student  $n$
- $D_i$  - difficulty of item  $i$
- $S_j$  - severity of rater  $j$
- $T_m$  - domain  $m$
- $F_k$  - difficulty of step up from category  $k-1$  to category  $k$

Jadual 3.1

*Tahap pengukuran, skala pengukuran dan penjelasan oleh Vygotsky (Dlm. Morrison, 2011, m.s. 135)*

Skor (Tahap Penguasaan)	Simbol	Kriteria
Kebolehan sangat tinggi	5	Kanak-kanak boleh menunjukkan respon tanpa bantuan
Kebolehan tinggi	4	Kanak-kanak boleh menunjukkan respon dengan hanya sedikit bantuan
Kebolehan sederhana	3	Kanak-kanak boleh menunjukkan respon dengan bantuan yang kerap
Kebolehan rendah	2	Kanak-kanak boleh menunjukkan respon hanya dengan bantuan
Kebolehan Sangat rendah	1	Kanak-kanak tiada respon

Model item respon teori (IRT) dikenal pasti oleh nombor bagi ciri-ciri item yang boleh dianggarkan. PRPF adalah model yang lebih daripada satu parameter yang mengukur kesukaran item semasa menganggarkan interaksi dengan kebolehan seseorang. Model Rasch boleh menyusun parameter kebolehan seseorang, parameter penilai, parameter domain dan parameter item dalam sistem pengukuran satu linear. Oleh kerana semua parameter berkongsi skala satu linear, maka semua bergabung untuk menghasilkan data pemerhatian yang diperlukan untuk menganalisis prestasi kanak-kanak.

Model Rasch mempunyai aspek kebebasan setempat (*local independence*) iaitu anggaran statistik mengukur secara bebas terhadap mana-mana penilai, calon (kanak-kanak) atau item (Linacre, 1994). Kekuatan menggunakan model ini ialah anggaran mengenai kebolehan calon adalah tidak bergantung kepada alat untuk menguji item, dan anggaran parameter item adalah bebas dan tidak bergantung kepada kumpulan calon (Crocker & Algina, 1986). Dalam erti kata lain, kebolehan kanak-kanak dan kesukaran item ditentukan tanpa menghiraukan latar belakang kanak-kanak yang ditaksir atau jenis pentaksiran yang digunakan selagi data yang diperolehi adalah fit (sesuai) dengan model.

Proses pentaksiran pembelajaran dan perkembangan kanak-kanak dalam kajian ini menyediakan pentaksiran merangkumi dua kohort iaitu kumpulan 1 (umur 2-3 tahun) dan kumpulan 2 (umur 3-4 tahun), kanak-kanak ditaksir secara pemerhatian berstruktur dan daripada hasil kerja merangkumi enam domain pembelajaran dan perkembangan dan dinilai oleh dua orang penilai menggunakan skala rating (1,2,3,4,5).

Model Rasch adalah model statistik dan kriteria penting menganalisis data adalah bukan kerana model fit dengan data, tetapi data adalah fit dengan model. Model Rasch mempunyai kemungkinan terdapat data yang tidak fit dengan model untuk menganalisis data dengan berjaya (Bond & Fox, 2001). *Misfitting* data boleh dipertimbangkan dan boleh disemak semula atau dibuang jika dapat mengesannya. Contohnya, rating penilai adalah tidak konsisten. Kejayaan menganalisa data ditentukan oleh kesesuaian statistik itu digunakan dalam setiap parameter. Analisis untuk fit statistik ditentukan jika data itu fit dengan model dan apa jangkauan pengukuran yang dilihat mampu pergi lebih jauh dan mendalam berbanding ukuran yang dijangka.

Kegunaan lain statistik ini adalah untuk menguji sejauh mana ciri terpendam dapat ditakrifkan, dan ia merujuk kepada kebolehpercayaan indeks pengasingan. Indeks ini menyediakan maklumat tentang sejauhmana elemen setiap faset terpisah untuk mentakrifkan kebolehpercayaan bagi setiap faset. kebolehpercayaan ini adalah kebolehpercayaan bersifat analog ke tradisional yang menggambarkan nisbah skor variation sebenar kepada skor variance yang diperhatikan.

Menggunakan PRPF, kanak-kanak, penilai, domain dan item tugas, ditentu ukur (*calibrated*) dengan berada di atas skala yang sama. Variable map membolehkan analisis untuk melihat semua faset atau komponen dalam satu gambarajah. Ia sangat berguna untuk melihat perbezaan komponen dalam satu skala dengan hanya satu unit pengukuran. Skala dalam variable map berbentuk unit log-odds atau logit. PRPF juga memperkenalkan statistik skala rating yang menyediakan maklumat bagaimana skala rating digunakan oleh penilai.

**Analisis Kajian Pengesahan.** Bagi peringkat pengesahan ini, penyelidik menguji ciri-ciri psikometrik instrumen IPHKT yang dibina menggunakan model pengukuran Rasch pelbagai facet (PRPF) dan Teori Generalizibiliti (Edu-G) untuk mentaksir perkembangan holistik kanak-kanak meliputi empat faset utama bagi kajian ini iaitu kanak-kanak, penilai, domain dan item yang akan dilihat dari segi kebolehpercayaan dan kesahan. Perbincangan analisis yang dijalankan bagi kajian ini diutarakan menurut persoalan kajian dua seperti di Bab 1.

**Kefungsian Kategori Skala.** Analisis terhadap kefungsian kategori skala bagi memastikan skala lima kategori yang digunakan adalah sesuai untuk mengukur perkembangan kanak-kanak. Perbincangan analisis yang dijalankan bagi kajian ini diutarakan menurut persoalan persoalan kajian seperti dalam Bab 1.

*Bagaimanakah kefungsian kelima-lima kategori skala yang digunakan bagi lima kategori (skala rating) dalam IPHKT?* Terdapat dua ciri yang penting dalam melihat frekuensi kategori. Yang pertama adalah bentuk taburan dan yang kedua adalah bilangan respons bagi setiap kategori. Taburan yang biasa seperti bentuk seragam, normal, bimodal dan taburan yang sedikit pencong adalah lebih baik daripada taburan yang tidak biasa seperti taburan yang sangat pencong (Linacre, 1999). Yang kedua adalah melihat bilangan frekuensi bagi setiap kategori. Kategori dengan frekuensi yang rendah adalah satu masalah kerana ianya tidak memberikan pemerhatian yang cukup untuk membuat anggaran yang stabil bagi nilai pemeringkatan. Nilai kategori yang mempunyai frekuensi yang sedikit adalah menunjukkan kategori tersebut tidak diperlukan ataupun merupakan kategori yang bertindihan dengan kategori yang lain. Nilai pemerhatian atau frekuensi yang paling minimum adalah sebanyak 10 pemerhatian (Linacre, 1999).

Satu lagi perkara yang dilihat bagi memastikan kefungsiian kategori adalah dengan melihat nilai min pengukuran. Nilai min pengukuran adalah berguna untuk dibuat penelitian bagi menentukan masalah awal dalam kategori skala. Min sampel didefinisikan sebagai min atau purata bagi anggaran keupayaan semua peserta bagi sampel yang dinilai dengan kategori respons yang tertentu, dengan min yang dikira merentasi ke semua pemerhatian dalam kategori tersebut. (Linacre, 1995). Sebagai contoh, sekiranya kategori (skor) 1, min ukurannya dicatatkan sebagai -2.5, nilai -2.5 boleh diinterpretasikan sebagai min keupayaan anggaran atau skor logit bagi individu-individu yang dinilai sebagai kategori 1 bagi mana-mana item dalam soal selidik. Nilai min ukuran ini dijangka akan meningkat nilainya apabila pembolehubahnya juga meningkat. Peningkatan yang sekata menunjukkan secara purata mereka yang mempunyai keupayaan yang tinggi atau kemahiran yang tinggi diberikan kategori yang tinggi, manakala bagi mereka yang berkeupayaan atau berkemahiran yang rendah diberikan kategori yang rendah. Sekiranya corak ini dicemari, maka ianya dianggap kurangnya kesekataan dalam pengukuran min dan adalah baik sekiranya dikurangkan penggunaan kategori dalam soal selidik.

*Adakah setiap kategori sepadan (fit) dengan model pengukuran Rasch pelbagai faset?* Selain penggunaan frekuensi kategori dan peningkatan secara sekata dalam min ukuran untuk menentukan kesesuaian skala kategori yang digunakan, nilai kategori fit statistik (Lopez 1996; Wright & Master 1982) juga digunakan. Nilai kesepadanan (*fit statistik*) memberikan satu lagi kriteria untuk menilai kualiti skala pengukuran. Menurut Linacre (1999), nilai *outfit* MNSQ yang melebihi 2 menunjukkan terdapat “*more misinformation than information*”, yang bermaksud bagi kategori tertentu terdapat banyak keadaan yang tidak sesuai dalam proses pengukuran.

Kategori yang begini memerlukan penyiataan lanjut secara empirikal dan merupakan calon yang sesuai untuk digabungkan dengan kategori yang berdekatan.

*Adakah nilai ambang (threshold) menunjukkan corak hierarki bagi skala pengukuran yang digunakan bagi setiap kategori (skala rating) dalam IPHKT? Nilai ambang (threshold) atau step calibrations adalah satu lagi kriteria penting yang perlu dilihat dalam menentukan kesesuaian penggunaan kategori bagi skala pengukuran yang digunakan. Nilai step calibrations adalah anggaran kesukaran bagi memilih kategori respons berbanding kategori yang lain. Sebagai contoh, bagaimanakah kesukaran memilih kategori “sangat berkebolehan” berbanding “berkebolehan”. Seperti nilai min ukuran, nilai step calibrations juga perlu meningkat secara sekata. Nilai yang tidak meningkat secara sekata merentasi skala pengukuran adalah dianggap tidak teratur.*

Magnitud bagi jarak antara nilai anggaran ambang (*threshold*) juga adalah penting. Jarak ambang adalah untuk menunjukkan setiap langkah mendefinisikan posisi yang berbeza bagi kategori. Ini menunjukkan bahawa nilai anggaran adalah terlalu dekat atau terlalu jauh bagi skala logit yang digunakan. Menurut Linacre (1999), nilai ambang perlu meningkat untuk menunjukkan perbezaan jarak antara kategori yang digunakan dan tidak melebihi 5 logit agar dapat mengelakkan jurang yang sangat besar di antara kategori yang digunakan.

Kaedah secara visual juga boleh digunakan untuk menyiasat perbezaan antara nilai ambang dengan melihat lengkung kebarangkalian yang menunjukkan kebarangkalian bagi satu kategori skala pengukuran. Setiap kategori seharusnya mempunyai puncak yang berbeza dalam lengkung kebarangkalian bagi menunjukkan setiap kategori respons yang paling mungkin bagi sebahagian pemboleh ubah yang diukur (Bond & Fox 2001). Kategori yang dilihat “mendatar” pada graf adalah

kategori yang tidak mungkin digunakan dan ini menunjukkan kategori tersebut tidak membantu dalam mendefinisikan perbezaan titik pada kategori (skor). Oleh itu, nilai ambang yang terlalu rapat atau tidak teratur dapat dilihat secara visual dengan bentuk lengkung kebarangkalian yang mendatar dianggap sebagai bermasalah.

**Faset Kanak-kanak.** Melihat kebolehpercayaan kanak-kanak, indeks pengasingan kanak-kanak dan nilai khi-kuasadua.

*Apakah nilai kebolehpercayaan kanak-kanak bagi IPHKT?* Nilai kebolehpercayaan bagi kanak-kanak dengan menggunakan model pengukuran Rasch adalah hampir sama dengan pengukuran yang dilakukan dengan menggunakan statistik pengukuran KR-20 ataupun Cronbach Alpha. Ianya menunjukkan sejauh mana sesuatu ukuran itu dapat dihasilkan semula. Nilai kebolehpercayaan yang dihasilkan daripada analisis Rasch adalah nilai kebolehpercayaan pengasingan (*seperation reliability*) yang memberikan maklumat bagaimana unsur-unsur dalam sesuatu faset dapat dipisahkan agar dapat mendefinisikan sesuatu faset dengan baik (Engelhard 2002; Myford & Wolfe 2003). Nilai kebolehpercayaan adalah nisbah di antara nilai varians “sebenar” bagi ukuran terhadap varians ujian. Secara matematik, pengiraan nilai kebolehpercayaan Rasch adalah seperti berikut:

$$R = \frac{SD^2}{RMSE^2 + SD^2}$$

di mana  $SD^2$  adalah ukuran varians “sebenar” dan  $RMSE$  adalah nilai punca kuasadua bagi ukuran ralat piawai. Nilai kebolehpercayaan Rasch ini adalah di antara 0 dan 1. Nilai yang tinggi menunjukkan tahap pengasingan yang lebih tinggi bagi unsur-unsur dalam faset tersebut (Wright & Masters 1982).

Adakah ukuran yang dilaksanakan dengan menggunakan IPHKT dapat mengasingkan perbezaan tahap keupayaan kanak-kanak? Untuk memastikan kanak-kanak dapat diasingkan kepada beberapa tahap yang berbeza terhadap pembolehubah yang diukur, ianya bergantung kepada kepelbagaian kumpulan kanak-kanak yang diukur. Walaupun pengukuran dilakukan terhadap kumpulan umur kanak-kanak seperti kumpulan umur dua, tiga dan empat, situasi pengukuran tetap memerlukan agar kanak-kanak dapat dipisahkan kepada beberapa tahap yang berbeza. Untuk melihat tahap keupayaan kanak-kanak, akan melihat pada nilai indeks pengasingan kanak-kanak (*seperation index*) dan nilai khi-kuasadua. Nilai indeks pengasingan adalah nilai sisihan piawai iaitu,

$$\text{Indeks pengasingan (seperation index), } G = \frac{SD}{RMSE}$$

di mana *SD* adalah nilai sisihan piawai sebenar dan *RMSE* adalah puncakuasa dua ralat. Tidak seperti nilai kebolehpercayaan yang berada di antara 0 sehingga 1, Fisher (1992) membincangkan perkara ini untuk susun atur indeks pengasingan dan nilai kebolehpercayaan. Selain itu, analisis *mean measure* menggunakan boxplot untuk melihat tahap kebolehan kanak-kanak mengikut umur 2 dan 3 serta 3 dan 4 tahun.

**Faset Penilai.** Analisis bagi faset penilai pula dilihat dari segi kebolehpercayaan antara penilai yang dilihat hasil daripada analisis peratus persetujuan penilai dan analisis nilai Rasch-Cohen Kappa. Ianya diikuti dengan melihat tingkh laku penilai dalam membuat penilaian. Penilai-penilai kemudiannya dilihat ukuran ketegasan dan diuji sama ada penilai mempunyai ketegasan yang berbeza dan konsistensi dalam membuat penilaian.



*Apakah nilai peratus persetujuan antara penilai?* Analisis dengan perisian FASETS juga menyediakan statistik bagi peratus persetujuan antara penilai. Nilai ini bermaksud sejauh mana penilai-penilai yang sama bersetuju dengan penilaian yang diberikan. Nilai peratus *observed* atau peratus persetujuan dari ujian melaporkan peratus persetujuan bagi penilaian yang dibuat di antara penilai. Peratus *expected* pula adalah peratus jangkaan yang melaporkan peratus yang sepatutnya berlaku sekiranya data sepadan dengan model pengukuran Rasch. Nilai yang baik yang biasa diterima menurut Linacre (2009a) adalah nilai yang melebihi sedikit daripada nilai peratus jangkaan.

*Apakah nilai kebolehpercayaan penilai?* Hasil daripada nilai peratus persetujuan yang dihasilkan oleh perisian FASETS juga boleh digunakan untuk mengukur nilai kebolehpercayaan iaitu statistik yang menunjukkan persetujuan antara penilai. Secara konsepnya, ianya adalah serupa dengan Cohen's Kappa dan merupakan indikator kepada kebolehpercayaan antara penilai. Rumus matematik untuk kebolehpercayaan penilai (Linacre 2009) adalah seperti berikut:

$$\text{Nilai kebolehpercayaan penilai} = \frac{(\% \text{ persetujuan ujian} - \% \text{ persetujuan jangkaan})}{(100 - \% \text{ persetujuan jangkaan})}$$

Nilai yang melebihi sifar adalah menunjukkan nilai persetujuan antara penilai yang tinggi dan nilai kurang daripada sifar menunjukkan penilai yang bersetuju adalah kurang daripada nilai jangkaan. Untuk menjadikan nilai kebolehpercayaan penilai menuruti julat konvensional di antara 0 hingga 1, rumus yang berikut digunakan, iaitu

$$\text{Cohen Kappa} = 1 - \frac{(\% \text{ persetujuan ujian} - \% \text{ persetujuan jangkaan})}{(100 - \% \text{ persetujuan jangkaan})}$$

Adakah guru penilai bertindak sebagai penilai yang bebas? Bagi menentukan kaedah tindakan penilai dalam menilai kanak-kanak, sekali lagi nilai peratus persetujuan antara penilai dilihat. Menurut Linacre (2008) nilai peratus persetujuan antara penilai (*observed*) yang dibuat adalah nilai peratus yang dipersetujui oleh seseorang penilai dengan penilai yang lain dalam menilai sesuatu item terhadap seseorang kanak-kanak. Nilai jangkaan (*expected*) melaporkan nilai persetujuan yang sepatutnya berdasarkan model pengukuran Rasch seperti yang dijelaskan di dalam Jadual 3.2.

Jadual 3.2

*Jadual perbandingan nilai ujian (observed) dan jangkaan (expected) oleh Linacre (2008)*

Peratus persetujuan penilai	Huraian
Ujian < Jangkaan (Observed < Expected)	Menunjukkan percanggahan pendapat, biasa berlaku bagi penilai yang tidak dilatih.
Ujian ~ Jangkaan (Observed ~ Expected)	Penilai bertindak secara tersendiri. Perlu disahkan dengan nilai <i>fit</i> statistik.
Ujian > Jangkaan (Observed somewhat < Expected)	Biasa bagi penilai yang terlatih. Latihan menitikberatkan persetujuan dengan yang lain tetapi <i>rating</i> memerlukan penilai bertindak secara bebas.
Ujian >> Jangkaan (Observed < Expected)	Penilai tidak bertindak secara bebas. Mungkin ada tekanan untuk bersetuju dengan yang lain.
Ujian > 90% (Observed > 90%)	Penilai adalah seperti " <i>rating machines</i> ". Sekiranya perkara ini berlaku, faset penilai perlu dikeluarkan dari model pengukuran

Apakah ukuran ketegasan penilai bagi IPHKT ? Adakah terdapat perbezaan ketegasan yang signifikan di antara penilai dalam menilai kanak-kanak? Adakah setiap guru menilai dengan konsisten dalam IPHKT? Bagi tiga persoalan kajian ni penyelidik menjawab dengan melihat hasil analisis FASETS berdasarkan model pengukuran Rasch terhadap ukuran ketegasan penilai serta nilai khi-kuasadua bagi menguji sama ada terdapat perbezaan ketegasan yang signifikan antara penilai. Nilai statistik kesepadanan pula digunakan untuk menentukan kesesuaian data penilai

membuat penilaian (Linacre, 2008). *Infit* adalah akronim bagi *information-weighted mean square residuals* yang mana lebih sensitif terhadap penilaian yang tidak dijangkakan terhadap seseorang pelajar, manakala *outfit* adalah akronim bagi *outfit sensitive fit statistics* di mana ia sensitif walaupun hanya terdapat satu nilai residuals (*observed – expected*) yang besar dan tidak dijangka. Nilai min kuasadua (MNSQ) dalam jadual ini adalah nilai MNSQ bagi *Infit* dan *outfit* dan nilai yang dijangkakan untuk membina pembolehubah dan membuat pengukuran (Wright & Masters 1982; Wright & Stone, 1979). Nilai yang boleh diterima menurut Linacre (2008) adalah di antara 0.5 dan 1.5. Penilai yang mempunyai nilai kurang daripada 0.5 ataupun melebihi daripada 1.5 adalah tidak konsisten dalam membuat penilaian.

**Faset Domain.** Adakah setiap domain IPHKT 1 dan 2 mempunyai ciri-ciri unidimensionaliti? Unidimensionaliti dalam analisis Rasch memerlukan sekurang-kurangnya pencapaian minimum 40 peratus *raw variance explained by measures* sebagai tanda ukuran instrumen unidimensionaliti yang baik, sebanyak 60 peratus.

**Faset Item.** Analisis item yang dijalankan bagi kajian ini adalah untuk menentukan kebolehpercayaan item dan kesahan. Nilai kebolehpercayaan item dilihat dari segi nilai indeks kebolehpercayaan dan indeks pengasingan. Kesahan yang dilihat di sini pula adalah kesahan kandungan dan kesahan konstruk. Kesahan kandungan dilihat dari nisbah kesahan kandungan seperti yang disarankan oleh Lawshe (1975). Bagi menentukan kesahan konstruk, menurut Wright dan Masters (1982) item-item dalam instrumen yang dibina mestilah dapat dipisahkan dengan sepatutnya secara statistik kepada beberapa tahap yang berbeza dan seterusnya menunjukkan satu arah sahaja, maka item-item dalam pembolehubah tersebut boleh diterima akal. Pola atau corak dalam penentuan item boleh memberikan satu gambaran jangkauan dan hierarki bagi sesuatu pembolehubah. Pola ini akan dibandingkan dengan kehendak

penyelidik untuk melihat sama ada ia mengesahkan pembolehubah yang dibina. Pada tahap ini, ia dapat mengesahkan kesahan konstruk bagi sesuatu pembolehubah (Wright & Masters 1982: 93).

*Apakah nilai kebolehpercayaan item dan indeks pengasingan item?* Nilai Kebolehpercayaan item menurut Bond & Fox (2001) adalah nilai yang menunjukkan sama ada item-item akan ada pada kedudukan yang sama sekiranya item tersebut diberi kepada sampel yang lain tetapi mempunyai keupayaan yang sama. Nilai kebolehpercayaan item dan indeks pengasingan item dilihat dari analisis PRPF dengan perisian FASETS. Nilai yang tinggi menunjukkan tahap pengasingan yang lebih tinggi bagi unsur-unsur dalam faset tersebut (Wright & Masters 1982).

*Di manakah kedudukan kesukaran item-item bagi setiap domain yang diukur?* Kedudukan kesukaran item-item dilihat melalui nilai penentukuran kesukaran item dalam nilai logit. Dengan melihat nilai ini, sebaran item-item dapat dipastikan sama ada mencukupi untuk mendefinisikan pelbagai tahap bagi konstruk yang diukur. Hanya item-item yang jelas bertabur dan terpisah dapat digunakan untuk menentukan arah bagi tujuan pengukuran. Untuk membandingkan kesukaran item-item dengan keupayaan kanak-kanak dan ketegasan penilai, penyelidik juga melihat rajah perbandingan ketiga-tiga faset tersebut dalam rajah pemetaan.

*Apakah ketepatan penentukuran yang dibuat?* Ketepatan penentukuran yang dibuat boleh dilihat dari nilai ralat piawai bagi setiap item. Nilai ralat piawai (S.E) yang rendah adalah nilai yang baik.

*Apakah respons item-item bagi setiap domain sepadan (fit) mengikut jangkaan model pengukuran Rasch?* Kesahan dalam model pengukuran Rasch dilihat dari nilai *fit* bagi setiap item. Item yang tidak *fit* adalah merupakan item yang perlu dimodifikasi atau dibuang. Item-item yang *fit* menunjukkan ia mengukur satu konstruk yang sama

(Purya, 2008). Nilai *fit* juga menunjukkan bahawa respons terhadap item-item adalah konsisten. Menurut Linacre (2002) nilai *fit* yang boleh diterima dalam pengukuran adalah antara 0.5-1.5. Dua nilai ukuran nilai *fit* yang dilihat dalam IPHKT adalah nilai *infit mnsq* dan nilai *outfit mnsq*.

*Adakah item-item bagi setiap pembolehubah bergerak pada satu arah? Untuk melihat sama ada pembolehubah bergerak pada satu arah yang sama menurut konstruk yang diukur, analisis polariti atau keselarian item dijalankan. Ukuran yang mempamerkan indeks positif bagi semua item menunjukkan semua item yang digunakan berfungsi ke arah yang selari untuk mengukur konstruk yang dibentuk. Menurut Bond & Fox (2001) polariti item atau *point measure correlation* adalah juga merupakan pengesanan awal kepada kesahan konstruk.*

*Bagaimanakah kedudukan kesukaran item-item hasil dari analisis berbanding dengan kesukaran item yang dijangkakan oleh pakar? Menurut Wright & Masters (1982), penentuan item-item akan meletakkannya pada titik-titik di sepanjang garis. Untuk memastikan item-item tersebut terpisah dengan sewajarnya, item-item tersebut mestilah tersusun menurut sepertimana yang telah ditetapkan oleh penyelidik dan pakar. Penyelidik telah meminta 15 orang dalam pendidikan awal kanak-kanak untuk menentukan kesukaran setiap item dalam IPHPT dan hasil analisis yang dilakukan adalah seperti dalam lampiran. Penyelidik telah menggunakan nilai median dalam menentukan kedudukan kesukaran item-item bagi setiap kemahirandalam IPHKT seperti yang disarankan oleh Wright dan Masters (1982). Kedudukan item-item seperti yang dijangkakan oleh kumpulan pakar mengikut nilai median, kemudiannya dibandingkan dengan kedudukan kesukaran item hasil daripada analisis Rasch dengan menggunakan perisian FASETS.*

Untuk membandingkan susunan oleh pakar dan hasil daripada analisis, satu rajah perbandingan di antara dua nilai tersebut dilakukan dengan melukis rajah plot sebaran (*scatter plot*) dengan menggunakan Microsoft Excel pada garisan Cartesan  $x$  dan  $y$ . Paksi- $x$  adalah mewakili nilai median bagi susunan item mengikut kesukaran yang dijangkakan oleh para pakar manakala paksi- $y$  adalah nilai kesukaran item hasil dari analisis dengan menggunakan perisian Facets Rasch. Penyelidik kemudiannya melukis satu garis pepenjur yang menunjukkan persetujuan bagi kedudukan item-item. Item-item yang tidak menunjukkan persetujuan akan berada jauh dari garis pepenjur yang digambarkan di dalam rajah melalui garis putus-putus.

**Dependability.** *Dependability* ujian dalam kajian ditentukan berdasarkan Teori Generalizabiliti (GT) menggunakan perisian *Education Generalizability* (EduG). *Dependability score* ujian akan menggunakan reka bentuk *Three faset partially Nested Random Design*. Analisis ini dilaksanakan menerusi empat langkah. Pertama, Model reka bentuk *Three faset partially Nested Random Design* yang mengandungi sumber varians yang akan dijelaskan oleh Analisis ANOVA iaitu menghuraikan lebih mendalam interaksi sumber varians menggunakan pengukuran sebenar (*absolute*). Kedua, kajian generalizabiliti (Kajian-G), dan, kajian keputusan (Kajian-D) dan ketiga nilai *optimization* bagi faset-faset.

*Apakah sumbangan setiap faset terhadap sumber varians mengikut Teori Generalizabiliti?* Daripada analisis, komponen varians yang menyumbang kepada *dependabilty* ujian akan ditunjukkan. Dalam kajian-G, Coef\_G relative ( $\hat{\epsilon}^2$ ) memaparkan tahap perbezaan ralat *relative* varian (*relative error variance*). Penggunaan jadual ANOVA digunakan untuk menganggarkan pekali kebolehpercayaan dan sumber ralat dalam model reka bentuk komponen sumber ralat. Mengikut GT, *universe* merujuk kepada keseluruhan keadaan populasi pengukuran (*measurement*

*condition*) dan populasi ini mengandungi pelbagai set kondisi pengukuran yang mungkin. Setiap set ini akan menjadi *faset*. Sebagai contoh, guru ingin mengukur prestasi kanak-kanak dalam domain perkembangan kreativiti. Maka reka bentuk kajian adalah *satu faset* iaitu  $n$  merupakan objek yang diukur. Objek pengukuran bagi kajian ini ialah kanak-kanak. Dalam contoh ini kemungkinan terdapat lima sumber utama kepada ralat iaitu ;

- i. Kanak-kanak ( $k$ )
- ii. item ( $i$ )
- iii. interaksi antara kanak-kanak dan item ( $kxi$ )
- iv. interaksi penilai dengan kanak-kanak dan item ( $p:ki$ )
- v. *residual* (ralat yang tidak diketahui) ( $e$ )

skor yang diperhatikan adalah termasuk interaksi keempat-empat sumber ralat dibenarkan. Secara formulanya;

$$X_{pkr} = p + v_p + v_i + v_{pj} + v_e$$

$$X_{pkr} = \text{skor diperhatikan}$$

$$M = \text{min populasi dan universe}$$

$$v = \text{kesan ralat untuk interaksi yang dinyatakan}$$

( $v_p, v_i, v_{pj}, v_e$ ) jelas dapat dilihat melalui kajian  $G$ , setiap sumber ralat yang dikenalpasti, dan dikira kesannya. Keadaan ini membenarkan pengiraan indeks kebolehpercayaan dalam sesuatu ujian atau pengukuran adalah lebih tepat dan lebih dipercayai.

*Apakah dependability score koefisien G bagi faset domain mengikut Teori Generalizabiliti? Kajian-G mengambil kira anggaran kesan setiap komponen varians yang secara relative terhadap skor responden. Anggaran komponen varians memberi refleksi terhadap magnitud ralat dalam mengeneralisasikan skor calon pada satu item*

kepada skor keseluruhannya (*universe score*). Anggaran komponen varian ini bukanlah ralat dalam generalisasi skor ujian.

*Apakah nilai optimization yang terbaik bagi faset kanak-kanak dan penilai menggunakan kajian D untuk meningkatkan koefisien G? Sementara kajian-D pula melibatkan kepelbagaian pengaplikasian reka bentuk yang telah ditentukan hasil manipulasi komponen varian dalam kajian-G. (Brennan, 1993; Shavelson & Webb, 1991). Analisis kajian G dan D akan menjawab dua sub-soalan kajian ini. Reka bentuk kajian-G adalah rekabentuk Two Faset (p:ki) Partially Nested Random Design. Dalam kajian-D Coef\_G absolute phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (absolute error variance). Perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai akan ditunjukkan.*

*Adakah terdapat interaksi bias (Differential Faset Functioning) penilai dan kumpulan umur kanak-kanak? Mengenal pasti interaksi bias antara penilai dan kumpulan umur kanak-kanak. Menurut Du et al. (1996), analisis DFF (Differential Faset Functioning) boleh dilaksanakan terhadap faset penilai, item, kanak-kanak tau faset yang lain yang mungkin memberi kesan kepada ketidakadilan dalam pengukuran kanak-kanak dalam penilaian berasaskan prestasi. Menurut Angoff (1993), DIF (Differential Item Functioning) adalah merujuk kepada pemeriksaan terhadap item-item yang mempunyai ciri-ciri statistik yang berbeza terhadap setting kumpulan yang berbeza (setelah mengambil kira perbezaan kebolehan mengikut kriteria tertentu). Menurut Du et al (1996) definisi yang sama dari segi prinsip boleh diaplikasikan bagi DFF. DFF boleh didefinisikan sebagai pemeriksaan terhadap item-item, penilai, kanak-kanak atau faset pengujian yang lain yang mempunyai ciri-ciri statistik yang berbeza terhadap setting kumpulan yang berbeza (setelah mengambil kira perbezaan*



kebolehan mengikut kriteria tertentu). Dalam kajian ini, sekali lagi model PRPF dengan perisian Facets Rasch digunakan untuk mengesan DFF.

Analisis DFF bagi kajian ini adalah untuk melihat kefungsiannya di antara faset penilai dan faset kumpulan umur kanak-kanak. Analisis ini menyediakan maklumat yang menunjukkan potensi sumber bagi berlakunya bias di dalam penilaian. Menurut McNamara (1996), kajian mengenai bias adalah kajian mengenai kesan interaksi dan juga dikenali sebagai interaksi bias (Myford & Wolfe, 2003).

Bagi menjawab persoalan kajian analisis bias telah dijalankan. Analisis bias dalam model PRPF adalah untuk menentukan sekiranya terdapat pola tertentu dalam pengukuran bias dan menentukan pengukuran yang dihasilkan berdasarkan kombinasi dua dan tiga faset (faset kanak-kanak dan faset item, faset kanak-kanak dan faset penilai serta faset kanak-kanak, faset penilai dan faset item) menyimpang dari apa yang dijangkakan berdasarkan model pengukuran Rasch. Dalam analisis bias, PRPF membandingkan nilai jangkaan dan nilai ujian dalam satu set data dan mengira perbezaan antara nilai jangkaan dan nilai ujian berada di antara julat yang normal.

Kajian mengenai bias adalah kajian mengenai kesan interaksi atau interaksi yang sistematik di antara faset tertentu. Analisis bias dalam PRPF biasanya digunakan untuk memeriksa pola ketegasan tingkah laku penilai dan kaitannya dengan kanak-kanak ataupun item tertentu (McNamara, 1996) dan juga ukuran bias bagi faset yang lain seperti kanak-kanak dan item. Sebagai contoh, bias akan terhasil apabila seorang penilai adalah lebih tegas dan lebih mudah, apabila memberi jangkaan bagi keupayaan kanak-kanak dan kesukaran item. Skor  $z$  dimodelkan sebagai taburan normal dan diperolehi dengan membahagikan nilai ukuran bias dengan ralat piawai. Nilai skor  $z$  yang kurang dari 2 atau lebih daripada 2 menunjukkan tingkah laku penilai yang tidak sepatutnya bagi kanak-kanak ataupun item.

## **Rumusan**

Metodologi kajian dalam bab ini dirangka untuk menguji instrumen pentaksiran prestasi, standard awal pembelajaran dan perkembangan awal kanak-kanak. Data yang dikumpul dapat dianalisis untuk mencapai objektif khusus kajian iaitu untuk menentukan indeks kebolehpercayaan, kesahan, ranking kanak-kanak, pengaruh penilai, bias penilai, kesukaran item mengikut domain, faset domain, pengukuhan antara penilai dan menesan sumber varians. Apabila semua proses dapat dijalankan dengan sistematik, instrumen ini dapat digunakan untuk mentaksir kanak-kanak.

University of Malaysia

## **Bab 4 Peringkat Rekabentuk Dan Pembinaan Instrumen**

### **Pengenalan**

Bahagian ini membincangkan tentang proses dan hasil bagi peringkat membangunkan instrumen, yang dimulakan dengan peringkat Rekabentuk dan pembinaan instrumen. Bab ini akan membincangkan peringkat rekabentuk terdiri daripada 5 langkah iaitu, mengenal pasti kumpulan sasaran, menentukan kandungan instrumen, mengenal pasti prestasi, menentukan skala dan mengenal pasti penilai. Peringkat pembinaan pula adalah terdiri daripada 5 langkah iaitu mengenal pasti standard, mengadaptasi instrumen, mentaksir instrumen, menentukan tujuan instrumen (membina instrumen) dan format instrumen (menguji, menyemak dan membangunkan manual instrumen).

### **Peringkat Rekabentuk (McNamara, 1999)**

*Menentukan Kumpulan Sasaran (Kanak-kanak).* Instrumen yang diadaptasi disesuaikan untuk kanak-kanak berumur 2 hingga 4 tahun. Untuk tujuan kajian ini, proses pengujian akan ditadbir kepada kanak-kanak di Taman Asuhan Kanak-kanak (TASKA) swasta.

*Menentukan Kandungan Instrumen.* Instrumen ini diadaptasi dengan tujuan untuk mentaksir dan menentukan kebolehan kanak-kanak berdasarkan prestasinya ke atas setiap item dalam domain serta menentukan item dan domain mana yang sukar bagi kanak-kanak kerana instrumen asal mengambil kira secara keseluruhan apa yang boleh dilakukan oleh kanak-kanak. Dengan mengambil kira faktor umur dan enam domain pembelajaran, penilaian menggunakan instrumen yang telah diadaptasi dapat menentukan tahap kesukaran item dan membezakan kebolehan kanak-kanak 2 hingga 4 tahun dan mematuhi standard. Kandungan instrumen ditentukan dengan mematuhi

standard penilaian kanak-kanak oleh Kementerian Pelajaran Malaysia. Menggunakan istilah domain perkembangan kajian seperti yang dibandingkan dalam Jadual 4.1 di bawah:

Jadual 4.1

*Perbandingan Istilah Domain Pembelajaran kanak-kanak*

No.	Tunjang Kurikulum Standard Prasekolah Kebangsaan (4,5,6 tahun)	Bidang Pembelajaran Kurikulum Kebangsaan Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan (0 hingga 4 tahun)	Setelah diadaptasi dan istilah yang digunakan dalam kajian ini (2 hingga 4 tahun)
1	Tunjang Komunikasi (Bahasa)	Perkembangan Bahasa, Komunikasi dan Literasi Awal (Bahasa)	Domain Bahasa
2	Tunjang Sains dan Teknologi	Perkembangan Pra-Matematik dan Pemikiran Logik	Domain Kognitif
3	Tunjang Kerohanian, Sikap dan Nilai (Rohani)	Perkembangan Sosio-emosi, Sahsiah dan Kerohanian (Rohani)	Domain Rohani
4	Tunjang Fizikal dan Estetika (Fizikal)	Perkembangan Fizikal dan Kemahiran Mengurus Diri (Fizikal)	Domain Fizikal
5	Tunjang Fizikal dan Estetika (Kreativiti)	Perkembangan Estetika dan Daya Kreatif (Kreativiti)	Domain Kreativiti
6	Tunjang Ketrampilan Diri (Sosioemosi)	Perkembangan Sosio-emosi, Sahsiah dan Kerohanian (Sosioemosi)	Domain Sosioemosi

**Mengenal pasti Prestasi.** Setiap domain dalam instrumen ini adalah berbentuk item prestasi di mana kanak-kanak akan menunjukkan kebolehannya melakukan tugas atau menyelesaikan masalah samada tanpa bantuan atau dengan bantuan.

**Menentukan Skala.** Penyelidik menentukan skor secara skala rating (skor 1,2,3,4 dan 5) yang boleh membezakan kebolehan kanak-kanak kerana kebolehan kanak-kanak melakukan tugas item) adalah berbeza samada dengan bantuan, tanpa bantuan atau berkebolehan melakukannya dengan sendiri. Skor yang diberikan mewakili peningkatan terhadap kebolehan kanak-kanak. Keupayaan mereka dirujuk kepada standard prestasi.

Standard prestasi merupakan satu set kriteria umum yang menunjukkan tahap-tahap prestasi yang perlu kanak-kanak pamerkan sebagai tanda bahawa sesuatu perkara itu telah dikuasai mereka. Ia memberi definisi “kejayaan” dan bagaimanakah hasil kerja dan tingkah laku yang dikatakan sangat berkebolehan. Tujuan diadakan standard prestasi ialah untuk mengenal pasti di manakah tahap penguasaan pengetahuan, kemahiran atau pengamalan nilai seseorang kanak-kanak itu. Standard prestasi memberi gambaran jelas atau panduan kepada guru tentang bagaimana hendak memberi skor kepada respon dan tindak balas prestasi oleh kanak-kanak. Standard prestasi dinyatakan dalam bentuk skor dan kriteria.

Kriteria merupakan pernyataan yang jelas tentang tahap perkembangan penguasaan kanak-kanak. Ia menjadi rujukan bagaimana prestasi kanak-kanak dinilai dan diberikan skor.

1. Cemerlang dan tanpa bimbingan (skor 5)
2. Sangat bagus dan sedikit bimbingan (skor 4)
3. Bagus dengan bimbingan (skor 3)
4. Masih memerlukan bimbingan (skor 2)
5. Sentiasa memerlukan bimbingan sepenuhnya (skor 1)

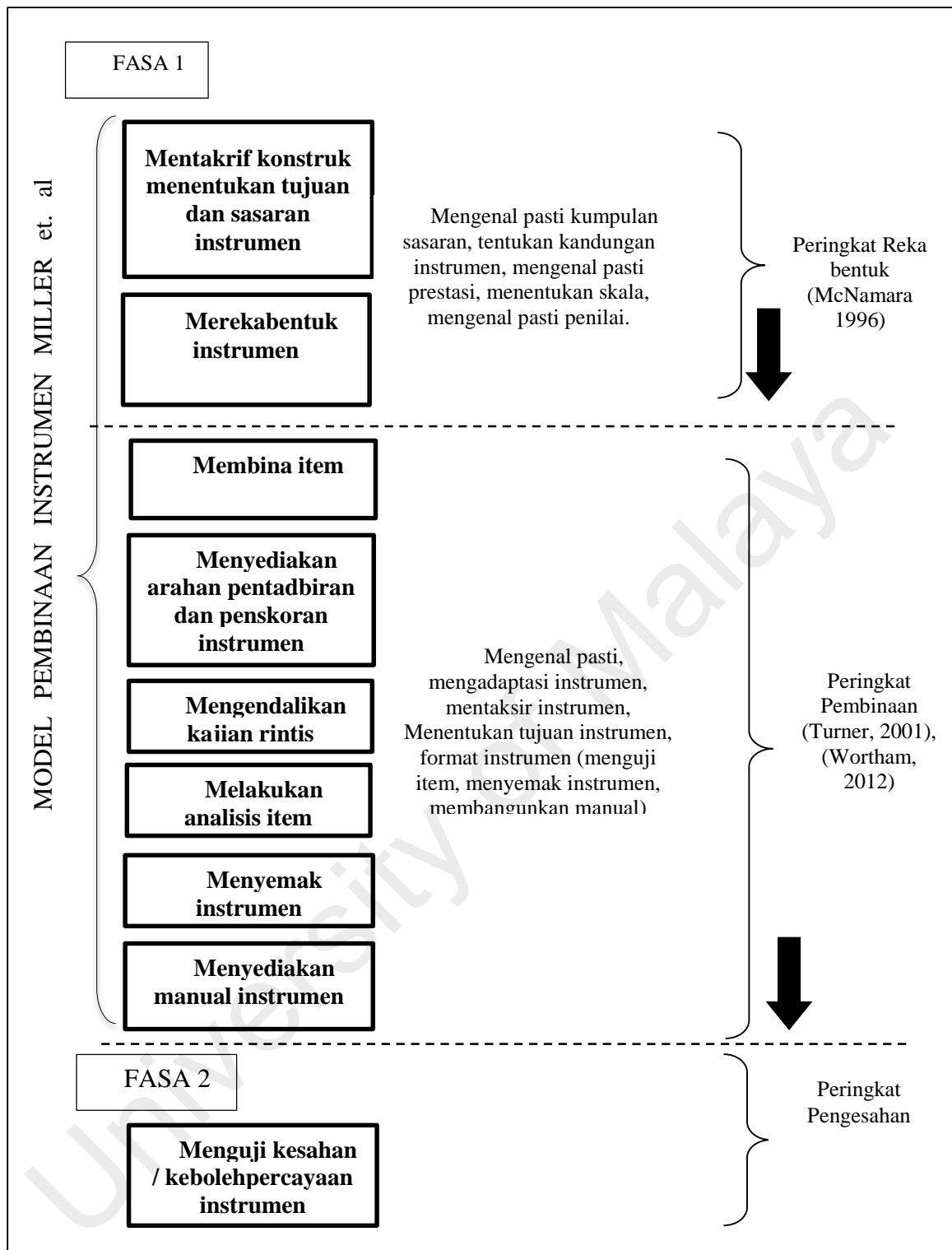
***Mengenal pasti penilai.*** Penyelidik merancang menggunakan pemerhatian oleh penilai kerana sampel kajian adalah kanak-kanak dan tidak boleh menilai diri mereka sendiri. Strategi pemarkahan secara keseluruhan bergantung kepada pemerhatian subjektif oleh penilai iaitu individu yang dilantik, diberi latihan khas dan melaksanakan proses pemerhatian terhadap sampel dan mencatat skor. Dua orang penilai akan memerhati sampel yang sama dan memberi skor dalam kajian rintis.

Dalam kajian sebenar, seramai 16 orang penilai dilatih dan menjalankan pentaksiran di taska. Pentaksiran dilakukan secara pemerhatian.

Menurut Merrill (1994), kaedah yang paling sesuai untuk menilai kemahiran pelajar melalui pemerhatian secara semulajadi. Menurut beliau lagi, kaedah ini dapat mengurangkan kebarangkalian dalam membuat kesilapan dalam andaian (*assumption*). Penyelidik juga mengambil kira beberapa perkara sepertimana yang disarankan oleh Carr & Claxton (2000) iaitu pentaksiran yang dilaksanakan mestilah boleh diuruskan oleh guru yang sibuk, memenuhi konstruk, fleksibel dan mempunyai kebolehpercayaan serta menyokong kurikulum. Claxton (2002) & Denton (2003) juga menyarankan agar pentaksiran adalah mudah untuk diuruskan oleh guru. Manakala Kelly (2001) menyatakan pentaksiran juga perlu ada keberkesanan kos. Salah satu kaedah untuk memastikan bahawa pentaksiran yang dilaksanakan adalah mudah untuk diuruskan ialah dengan menjalankannya semasa proses pengajaran dan pembelajaran. Oleh yang demikian, mengambil kira penglibatan guru sebagai penilai dapat memberikan akauntabiliti yang lebih seterusnya memberikan pendedahan baru aspek pentaksiran kanak-kanak.

### **Kerangka Pembinaan Instrumen**

Rajah 4.1 menunjukkan kerangka pembinaan instrumen dengan sepuluh langkah yang digabungkan oleh McNamara, 1996; Turner, 2001 dan Wortham, 2012 berdasarkan Model Miller et al. (2013).



Rajah 4.1. Kerangka Pembinaan Instrumen

## Peringkat Pembinaan Instrumen (Turner, 2001; Wortham, 2012)

**Mengenal pasti Standard.** Standard Pentaksiran Perkembangan Murid KPM dirujuk untuk melakukan penambahan ke dalam instrumen. Isi bagi ruang pertambahan di rujuk dalam Lampiran Instrumen IPHKT. Perkara yang ditambah adalah Fokus, Standard Kandungan dan Standard Pembelajaran manakala item prestasi dan aktiviti adalah daripada instrumen asal, seperti dalam Rajah 4.2 berikut:

Kod	Fokus	Standard Kandungan	Standard Pembelajaran	Item prestasi	Aktiviti/ Soalan	Catatan
-----	-------	--------------------	-----------------------	---------------	------------------	---------

*Rajah 4.2.* Penambahan aspek baru pentaksiran dalam rekabentuk instrumen

Fokus adalah bidang pembelajaran yang ingin dikembangkan dalam diri kanak-kanak. Fokus dinyatakan melalui Standard Kandungan. Standard Kandungan merupakan pernyataan spesifik tentang perkara yang kanak-kanak patut tahu dan boleh lakukan. Setiap Standard Kandungan diperincikan lagi kepada Standard pembelajaran. Standard Pembelajaran merupakan satu penetapan kriteria atau indikator pencapaian yang boleh diukur, untuk memastikan penguasaan setiap Standard Kandungan.

**Mengadaptasi Instrumen.** Instrumen ini menggambarkan keseluruhan kemahiran pada tahap umur dua hingga empat tahun yang memerlukan kanak-kanak berhubung dengan domain-domain pembelajaran dan perkembangan. Tahap umur dan instrumen dibahagikan kepada dua iaitu Kumpulan 1 (2 hingga 3 tahun) dan Kumpulan 2 (3 hingga 4 tahun), yang setiap satu mengandungi item prestasi yang memerlukan kanak-kanak melakukan sesuatu tugas. Aktiviti yang disediakan akan menterjemahkan item prestasi tersebut. Setiap instrumen mengandungi enam domain dan mempunyai pernyataan item prestasi yang khusus yang merangsang tindakan kanak-kanak. Domain tersebut adalah Domain Kognitif, Domain Bahasa, Domain Kreativiti, Domain Fizikal, Domain Sosioemosi dan Domain Rohani. Instrumen yang telah diadaptasi diringkaskan sebagai Instrumen Pentaksiran Holistik Kanak-kanak



Taska (IPHKT). IPHKT 1 untuk kanak-kanak berumur 2 dan 3 tahun, manakala IPHKT 2 untuk kanak-kanak berumur 3 dan 4 tahun. Instrumen boleh dirujuk di dalam Lampiran L dan M.

Setiap domain mengandungi jumlah item yang berbeza. Rekabentuk instrumen setelah diadaptasi adalah mengandungi Kohort, Domain, Instrumen, Kod, Fokus, Standard Kandungan, Standard Pembelajaran, Item Prestasi, Aktiviti dan Catatan seperti Rajah 4.3 berikut:

Kohort ( ) – Domain ( ) - Instrumen _____						
Kod.	Fokus	Standard Kandungan	Standard Pembelajaran	Item Prestasi	Aktiviti/	Catatan

Rajah 4.3.. Reka bentuk instrumen pentaksiran mengikut domain

	Domain	Nama Domain	Kod
Kump. 1 (2-3 tahun)	1	Kognitif	K1-K21
	2	Bahasa	B1-B16
	3	Kreativiti	KR1-KR21
	4	Fizikal	F1-F39
	5	Sosioemosi	S1-S61
	6	Rohani	R1-R10
		Jumlah item	171

Rajah 4.4. Jumlah item bagi semua domain dalam IPHKT 1

	Domain	Instrumen	Kod
Kump. 2 (3-4 tahun)	1	Kognitif	K1-K37
	2	Bahasa	B1-B25
	3	Kreativiti	KR1-KR21
	4	Fizikal	F1-F29
	5	Sosioemosi	S1-S62
	6	Rohani	R1-R9
		Jumlah item	181

Rajah 4.5. Jumlah item bagi semua domain dalam IPHKT 2

Berdasarkan tindakan atau prestasi yang ditunjuk oleh kanak-kanak, skor akan diberikan yang menunjukkan sejauhmana tahap kebolehannya. Setiap kanak-kanak akan ditaksir oleh dua orang penilai melalui pemerhatian secara langsung dan hasil kerja. Penilai 1 dan 2 akan menggunakan Instrumen Pentaksiran Domain Kanak-

kanak (PDK) bagi setiap enam domain (lihat lampiran B-borang pentaksiran). Borang PDK mengandungi maklumat kriteria asas pentaksiran, item prestasi, skala 1 hingga 5. Kanak-kanak hanya akan menerima skor rating sahaja daripada penilai, tiada grad atau ranking dicatat dalam borang. Pentaksiran Kumpulan 1 dijalankan selama dua hari dan Kumpulan 2 selama dua hari juga secara berasingan.

**Mentaksir Instrumen. *Membezakan tahap kebolehan kanak-kanak.*** Antara kanak-kanak dalam kumpulan umur 2 dan 3 tahun, atau 3 dan empat tahun, terdapat kanak-kanak yang berkebolehan tinggi dan rendah melakukan tugas (item). Kedudukan ini yang perlu ditentukan kerana berdasarkan definisi operasi kebolehan kanak-kanak dalam kajian ini, terdapat kanak-kanak yang mempunyai prestasi *above-level*. Oleh itu, pengesanan terhadap kebolehan mereka boleh ditaksir dan disahkan berdasarkan nilai ukuran dalam kajian ini.

***Membezakan ketegasan penilai semasa mentaksir.*** Penilai adalah individu yang memberi skor. Maka, penilai yang memiliki ketegasan dan kelembutan (*severity and leniency*) perlu ditentukan. Penilai turut menilai prestasi kanak-kanak terhadap domain-domain instrumen, samada penilai ini mempunyai bias atau tidak, maka indeks bias antara penilai dan domain perlu ditentukan.

***Menentukan Tahap Kesukaran Item.*** Pencapaian kanak-kanak ditentukan melalui prestasi yang ditunjukkan semasa melakukan aktiviti atau tugas (item). Terdapat item yang sukar dan mudah dalam domain. Maka, item yang sukar dan mudah bagi kanak-kanak perlu ditentukan berdasarkan nilai ukuran. Kedudukan item akan disusun mengikut tahap kesukaran menggunakan nilai logit.

***Tujuan Membina Instrumen.*** Setiap instrumen yang standard harus mempunyai tujuan yang jelas. Penerangan bagi tujuan instrumen adalah untuk

merangka pembentukan instrumen untuk mengukur apa yang ingin diukur. Ia juga membolehkan evolusi instrumen apabila reka bentuk dan pembinaan instrumen yang terbaru dilakukan. Semasa membina instrumen, harus menjelaskan apa perkembangan atau ciri-ciri instrumen yang akan dilaksanakan, bagaimana keputusan yang akan digunakan, dan siapa yang akan mengambil ujian itu. Instrumen yang disediakan untuk kanak-kanak sangat berbeza daripada ujian yang direka untuk orang dewasa.

Pentaksiran biasanya mempunyai tiga tujuan iaitu untuk menggalakkan pembelajaran, mengukur pencapaian individu dan untuk menilai program (Airasian 1994; Pellegrino, 2001). Menurut Foster (2005) pentaksiran adalah untuk pembelajaran dan digunakan untuk melaporkan hasil pembelajaran. Pentaksiran bagi pembelajaran biasanya digunakan untuk menggambarkan pentaksiran yang membawa kepada laporan pencapaian atau prestasi. Ia biasanya dikaitkan dengan pengukuran pencapaian individu. Pentaksiran untuk pembelajaran biasanya dikaitkan dengan galakan bagi pembelajaran. Manakala pentaksiran semasa pembelajaran menggambarkan pembentukan tugas pentaksiran yang direkabentuk untuk melibatkan pelajar dalam aktiviti yang mencabar dan bermakna dalam pembelajaran, iaitu mengintegrasikan pembelajaran dan pentaksiran. Pentaksiran yang baik boleh direkabentuk untuk menggalakkan pembelajaran dan pada masa yang sama digunakan untuk melaporkan hasil pembelajaran (Stobart, 2004).

Bagi kajian ini, tujuan pembinaan instrumen ini adalah untuk melaporkan prestasi kanak-kanak dalam domain-domain perkembangan secara holistik kepada mana-mana pihak yang berkeperluan yang inginkan maklumat berkaitan (pentaksiran kanak-kanak), menaikkan peranan guru sebagai penilai yang berautoriti untuk menilai perkembangan kanak-kanak secara terlatih. Pentaksiran yang dilakukan semasa pembelajaran dan pengajaran kanak-kanak dapat menyokong perkembangan

pembelajaran kanak-kanak dan dapat menjadi penggalak untuk penguasaan pembelajaran masteri serta menggunakan penilaian untuk meningkatkan kredibiliti guru penilai menjadi lebih berkeyakinan dalam melaksanakan tugas penilaian.

*Menentukan format instrumen.* Format ujian adalah berdasarkan kepada penentuan yang dibuat mengenai tujuan penilaian dan ciri-ciri pengambil ujian. Hasil keputusan format instrumen akan mempengaruhi bagaimana pentadbir instrumen menjalankan pengujian item dan mengambil kira bagaimana respon dan tindak balas pengambil ujian (Brown, 1983). Salah satu pertimbangan adalah sama ada ujian akan menjadi lisan atau bertulis. Penilaian yang direka untuk kanak-kanak kecil biasanya disampaikan secara lisan oleh pentadbir instrumen. Satu alternatif adalah dengan menggunakan tindak balas psikomotor; kanak-kanak itu diberikan objek untuk memanipulasi atau diminta untuk melaksanakan tugas yang fizikal.

Dalam kajian ini, format instrumen direka berdasarkan prosedur yang telah ditetapkan dalam Modul Pentaksiran Perkembangan Kanak-kanak yang dibina oleh Bahagian Pembangunan Kurikulum, Kementerian Pelajaran Malaysia pada tahun 2010. Prosedur ini diambil kira kerana modul pentaksiran ini adalah satu-satunya panduan yang dihasilkan mengikut budaya kanak-kanak Malaysia dan ia merupakan sistem yang telah dilaksanakan untuk mentaksir kanak-kanak di Malaysia. Format instrumen boleh rujuk lebih jelas di dalam Lampiran

### **Dapatan Kajian Rintis**

Kajian rintis dilaksanakan untuk melihat kesesuaian instrumen pentaksiran dengan model pengukuran yang digunakan dalam kajian ini. Setelah instrumen diadaptasi dan direka bentuk semula, kajian rintis dijalankan dan mendapatkan nilai kebolehpercayaan bagi faset kanak-kanak, faset penilai, faset domain dan faset item. Kesahan konstruk juga diperolehi berdasarkan dapatan kajian rintis yang akan

membuktikan konstruk dalam instrumen ini adalah sah dan membuktikan jumlah item dan sampel yang mencukupi. Beberapa perkara pembuktian yang diperolehi daripada dapatan kajian rintis ini.

**Nilai Kebolehpercayaan.** Nilai kebolehpercayaan kajian ini ialah bagi kanak-kanak (0.99), penilai(0.97), dan domain perkembangan (1.00). Jadual 4.2 menunjukkan nilai kebolehpercayaan kanak-kanak, penilai dan domain perkembangan bagi kajian ini adalah di antara 0.97 hingga 1.00 dengan indeks pengasingan di antara 6.02 hingga 23.30. Nilai ini adalah memenuhi kehendak model pengukuran Rasch. Nilai kebolehpercayaan responden  $\geq 0.8$  dan nilai indeks pengasingan  $\geq 2$  adalah indeks yang diterima baik.

Jadual 4.2

*Nilai kebolehpercayaan dan indeks pengasingan bagi 4 faset Instrumen bagi kajian rintis*

Faset	Nilai Kebolehpercayaan	Indeks Pengasingan
Kanak-kanak	0.99	9.81
Penilai	0.97	6.02
Domain	1.00	23.30

**Kesahan Konstruk.** Melihat nilai kesahan konstruk setiap faset berikut merujuk kepada nilai Infit MNSQ dan Outfit MNSQ dalam lingkungan 0.6 hingga 1.4. Julat statistik Infit MNSQ yang diterima seperti yang dilaporkan dalam kajian lepas ialah 0.8 hingga 1.2 (Wright & Linacre, 1984), 0.5 hingga 2.0 (Myford & Mislevy, 1995), dan 0.75 hingga 1.3 (McNamara, 1996). Statistik Infit MNSQ dan Outfit MNSQ dengan julat 0.5 hingga 1.5 adalah nilai julat yang sering digunakan untuk skala rating analisis Rasch dan akan digunakan dalam kajian ini (Linacre, 2014). Bagi statistik Outfit MNSQ pula, sama dengan Outfit MNSQ tetapi lebih sensitif terhadap outliers. Julat yang diterima untuk mengenal pasti misfitting Outfit MNSQ adalah sama dengan Infit MNSQ iaitu 0.5 hingga 1.5. Kedua-dua statistik Infit MNSQ dan Outfit MNSQ

disemak untuk mengenal pasti data yang tidak fit (sesuai) dengan PRPF. Nilai kesahan kajian ini akan dinyatakan dalam huraian hasil dapatan setiap faset.

Secara keseluruhannya, nilai kesahan merujuk kepada nilai min kuasadua (MNSQ) adalah nilai MNSQ bagi Infit dan outfit dan nilai yang dijangkakan untuk membina pembolehubah dan membuat pengukuran (Wright & Masters 1982; Wright & Stone, 1979). Nilai yang boleh diterima menurut Linacre (2014) adalah di antara 0.6 dan 1.4. Penilai yang mempunyai nilai kurang daripada 0.6 ataupun melebihi daripada 1.4 adalah tidak konsisten dalam membuat penilaian. Dapatan telah menunjukkan penilai menilai dengan ketegasan yang berbeza. Nilai MNSQ bagi Infit dan outfit berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai setiap kanak-kanak secara konsisten seterusnya menunjukkan data penilaian mereka mempunyai nilai kesahan. Begitu juga dapatan yang ditunjukkan oleh kanak-kanak dan domain perkembangan menunjukkan bacaan dalam julat yang sepatutnya seterusnya menjadikan data itu sah dan dipercayai.

**Nilai Ukuran dan kesahan kebolehan kanak-kanak.** Jadual 4.3 menunjukkan laporan ukuran kebolehan kanak-kanak tiga kebolehan tinggi dan 3 kebolehan rendah iaitu nilai ukuran kanak-kanak, nilai RMSE, ralat piawai (SE), nilai indeks pengasingan dan nilai khikuasa dua. Bagi nilai ukuran kanak-kanak adalah di antara -0.57 (ralat piawai= 0.06, kanak-kanak 2) iaitu ukuran kanak-kanak yang terendah dan 0.96 (ralat piawai = 0.08, kanak-kanak 49) iaitu ukuran kanak-kanak yang paling tinggi. Nilai RMSE bagi faset kanak-kanak adalah 0.07. Nilai indeks pengasingan adalah 9.81 menunjukkan kanak-kanak ini dapat dipisah atau diasingkan kepada keupayaan yang berbeza. Ini disahkan lagi dengan nilai khikuasadua yang signifikan di mana bagi nilai  $\chi^2 = 5759.5$ ,  $p < 0.01$ ,  $df = 59$ . Nilai indeks pengasingan

dan nilai khikuasa dua yang signifikan juga menunjukkan bahawa ukuran kanak-kanak yang diperolehi dapat memisahkan kanak-kanak kepada keupayaan yang berbeza.

Secara matematik, pengiraan nilai kebolehpercayaan Rasch adalah seperti berikut:

$$R = \frac{SD^2}{RMSE^2 + SD^2}$$

di mana  $SD^2$  adalah ukuran varians “sebenar” dan RMSE adalah nilai punca kuasadua bagi ukuran ralat piawai. Nilai kebolehpercayaan Rasch ini adalah di antara 0 dan 1. Nilai yang tinggi menunjukkan tahap pengasingan yang lebih tinggi bagi unsur-unsur dalam faset tersebut (Wright & Masters, 1982). Dapatan menunjukkan semua faset mempunyai nilai kebolehpercayaan yang tinggi.

Nilai kesahan kebolehan kanak-kanak ini merujuk statistik Infit MNSQ dan Outfit MNSQ yang disamak untuk mengenal pasti data yang serasi dengan model pengukuran ini. Hasil dapatan menunjukkan bacaan nilai Infit MNSQ (0.89 – 1.28) dan Outfit MNSQ (0.85 – 1.30) adalah dalam julat yang diterima dan menunjukkan data faset kanak-kanak memenuhi jangkaan model pengukuran ini seterusnya mempunyai nilai kesahan.

Jadual 4.3  
*Kedudukan kebolehan kanak-kanak bagi kajian rintis*

Kanak-kanak	Nilai ukuran kebolehan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
49	0.96	0.08	0.92	0.85
47	0.17	0.08	0.89	0.85
55	0.15	0.08	0.90	0.85
3	-0.19	0.06	1.28	1.30
1	-0.33	0.06	1.10	1.09
2	-0.57	0.06	1.08	1.10

*Nota.* Nilai khi-kuasadua: 5759.5 ; df: 59; Signifikan (kebarangkalian) : 0.00 Nilai RMSE: 0.07; Nilai Pengasingan Indeks: 9.81

Jadual 4.4 menunjukkan statistik pengukuran faset penilai terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.07 logit (SE= 0.01) bagi Penilai 2 iaitu penilai yang pemurah (*lenient*) sehingga 0.07 logit (SE= 0.01) bagi penilai Penilai 1 iaitu penilai yang tegas. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan.

Hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 74.5$ ,  $df = 1$ ,  $p < 0.01$ ) antara penilai. Ini menunjukkan bahawa penilai ini menilai dengan ketegasan yang berbeza.

Nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai perkembangan kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Nilai Infit dan Outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.6 dan 1.4. Iaitu Penilai 1 (Infit MNSQ: 0.99, Outfit MNSQ: 0.98) dan Penilai 2 (Infit MNSQ: 1.04, Outfit MNSQ: 1.03) Bacaan berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai setiap kanak-kanak secara konsisten seterusnya menunjukkan data penilaian mereka mempunyai nilai kesahan.

Jadual 4.4

*Nilai Ukuran Perbezaan Ketegasan Penilai bagi Kajian Rintis*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
Penilai 1	0.07	0.01	0.99	0.98
Penilai 2	-0.07	0.01	1.04	1.03

*Nota.* Nilai khi-kuasadua: 74.5 ; df: 1; Signifikan (kebarangkalian) : 0.00

#### **Nilai Ukuran Domain dan Kedudukan mengikut Tahap Kesukaran.**

Berdasarkan Jadual 4.5, domain rohani pada 0.96 logit (SE = 0.03) adalah domain yang paling sukar, manakala domain fizikal pada -0.84 logit (SE = 0.02) adalah item yang paling mudah. Nilai khi-kuasadua adalah 3035.1, df adalah 5,  $p < 0.01$  bagi



domain menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara domain-domain. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara domain-domain.

Ketepatan pengukuran dilihat dari nilai ralat piawai (SE) dalam Jadual 4.5. Nilai ralat piawai bagi semua domain berada di antara 0.01 hingga 0.03. Jadual 4.5 menunjukkan julat nilai Infit MNSQ semua domain berada di antara 0.81 – 1.44 manakala nilai Outfit MNSQ berada di antara 0.84 – 1.45. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2014). Rajah 4.6 menunjukkan kedudukan domain mengikut kesukaran. Domain yang paling atas adalah paling sukar iaitu Domain Rohani, manakala yang paling bawah adalah yang paling mudah iaitu Domain Fizikal.

Jadual 4.5

*Nilai Ukuran dan Kedudukan Domain Perkembangan (Susah ke mudah)*

Domain	Nilai ukuran	Logit SE	Infit MNSQ	Outfit MNSQ
Rohani	0.96	0.03	1.44	1.45
Bahasa	0.17	0.03	1.20	1.11
Sosioemosi	0.15	0.01	0.96	0.98
Kreativiti	-0.19	0.02	0.84	0.88
Kognitif	-0.25	0.02	0.81	0.84
Fizikal	-0.84	0.02	1.14	1.07

*Nota.* Nilai khi-kuasadua : 3035.1 ; df: 5 ; Signifikan (kebarangkalian) : 0.00

Dapatan menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara domain-domain. Selain itu, ketepatan pengukuran dilihat daripada nilai ralat piawai (SE) berdasarkan Jadual 4.5 ini.

logit	+Kanak-kanak	-Penilai	-Domain	-Item	Skala
3	**				(5)
2	** * ** **** ***** ** **** *** **			*	4
1	** **** *** * **** **** *** ** ** ** ** **		Rohani	*	
0	** * * ** * *	Penilai 1 Penilai 2	Bahasa Sosioemosi Kreativiti Kognitif	*** **** ** ** ***** *** ***** ***** *** *** **** *	3
-1			Fizikal		(1)
Ukuran	* = 1	-Penilai	-Domain	* = 1	Skala

Rajah 4.6. Kedudukan tahap kesukaran domain dan ketegasan penilai

Berdasarkan hasil dapatan analisis pertama, Nilai kebolehpercayaan bagi faset kanak-kanak, penilai dan domain perkembangan dengan menggunakan model pengukuran Rasch adalah hampir sama dengan pengukuran yang dilakukan dengan menggunakan statistik pengukuran KR-20 ataupun Cronbach Alpha. Ianya menunjukkan sejauh mana sesuatu ukuran itu dapat dihasilkan semula.

Nilai kebolehpercayaan yang dihasilkan daripada analisis Rasch adalah nilai kebolehpercayaan pengasingan (separation reliability) yang memberikan maklumat bagaimana unsur-unsur dalam sesuatu faset dapat dipisahkan agar dapat mendefinisikan sesuatu faset dengan baik (Engelhard 2002; Myford & Wolfe 2003). Nilai kebolehpercayaan adalah nisbah di antara nilai varians “sebenar” bagi ukuran terhadap varians ujian.

## **Latihan Guru**

**Prosedur Pengumpulan Data Latihan.** Sampel kajian ini dipilih secara rawak daripada taska swasta. Instrumen pemerhatian yang mengandungi 62 item digunakan dan ditadbir kepada enam orang kanak-kanak berumur dua tahun dan tiga tahun. Rakaman video terhadap enam orang kanak-kanak ini dibuat terlebih dahulu.

Pengumpulan data melibatkan dua orang guru taska yang bertindak sebagai penilai yang telah menjalani bengkel latihan dan menilai respons kanak-kanak melalui tayangan video. Strategi pemarkahan secara keseluruhan bergantung kepada pemerhatian oleh penilai iaitu individu yang dilantik, diberi latihan khas dan melaksanakan proses pemerhatian terhadap sampel dan mencatat skor. Instrumen latihan boleh dirujuk dalam lampiran

**Jadual Penilaian.** Para penilai yang dipilih adalah berdasarkan jadual penilaian yang dibina bagi memenuhi Model Pengukuran Rasch Pelbagai Faset (PRPF). Menurut Linacre (1994,1997), untuk menghasilkan satu pengukuran yang baik, perlu ada perkaitan atau hubungan antara setiap faset bagi setiap parameter agar penilaian dapat ditentukan tanpa keraguan dalam satu rangka rujukan. Satu keperluan yang sangat penting dalam membuat pelan penilaian adalah mempunyai pertalian yang mencukupi di antara semua elemen bagi semua faset di mana setiap parameter boleh

dianggarkan dalam satu rangka rujukan. Jadual penilaian perlu memenuhi keperluan pertalian atau pertindihan agar setiap elemen yang boleh dibandingkan dengan jelas dengan setiap elemen. Ini akan menghasilkan pengukuran yang tepat bagi setiap parameter dalam satu rangka rujukan. Jadual 4.6 menunjukkan jadual penilaian tiga orang penilai yang menilai enam orang kanak-kanak bagi enam domain perkembangan. Ini menunjukkan semua penilai menilai kanak-kanak yang sama dan domain yang sama.

Jadual 4.6  
Jadual Penilaian

Kanak-kanak	Penilai		
	8	13	18
1	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R
2	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R
3	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R
11	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R
12	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R
13	K/B/KR/F/S/R	K/B/KR/F/S/R	K/B/KR/F/S/R

*Nota.* K: Kognitif; B: Bahasa; KR: Kreativiti; F: Fizikal; S: Sosioemosi; R: Rohani

Bagi memastikan para penilai memahami proses penilaian yang perlu dilakukan, bengkel latihan telah dijalankan kepada para penilai sebanyak empat kali. Bengkel latihan yang dijalankan meliputi tayangan video kanak-kanak, kefahaman item kepada penilai. Perbincangan skor berasaskan rubrik serta panduan kepada penilai semasa membuat pemerhatian kepada prestasi kanak-kanak. Penilai diberi masa selama 6 bulan untuk membuat pemerhatian kepada perkembangan awal kanak-kanak yang dijadikan sampel kajian dan pada bulan keenam barulah bengkel latihan diberikan dan penilaian diberikan menggunakan instrumen. Penilaian akhir tahun

melihat secara keseluruhan perkembangan kanak-kanak. Ini adalah kerana penilai yang terlibat terdiri daripada guru taska dan telah membuat pemerhatian dalam tempoh yang telah ditetapkan.

### **Dapatan dan Perbincangan Kajian Data Latihan**

Analisis terhadap kefungsian kategori skala bagi memastikan skala lima kategori yang digunakan adalah sesuai untuk mengukur perkembangan kanak-kanak. Seterusnya analisis dijalankan terhadap empat faset utama kajian ini iaitu kanak-kanak, penilai, domain dan item yang akan dilihat dari segi kebolehpercayaan dan kesahan.

Dalam kajian, lima kategori skala digunakan dalam instrumen latihan bagi kesemua konstruk dalam instrumen. Kanak-kanak yang sentiasa menampakkan ciri-ciri perkembangan yang dikaji dianggap mempunyai tahap pengukuran yang sangat tinggi dan ditandai dengan nilai skala pengukuran 5, manakala kanak-kanak yang kurang menampakkan ciri-ciri perkembangan yang dikaji dianggap mempunyai tahap perkembangan kebolehan yang sangat rendah dan ditandai dengan nilai skala pengukurannya 1.

Jadual 4.7 berikut menunjukkan bilangan pemerhatian dan min ukuran bagi skala pengukuran lima kategori bagi enam domain perkembangan instrumen. Daripada bilangan pemerhatian atau frekuensi, bagi setiap kategori mempunyai lebih daripada sepuluh respons. Nilai min ukuran bagi kategori 1 adalah -1.71, bermakna min anggaran bagi kanak-kanak yang dinilai bagi kategori 1 merentasi item-item domain perkembangan adalah -1.71 logit. Bagi kanak-kanak yang diberi 2 bagi mana-mana item, min anggaran adalah -0.70. Dapat dilihat di sini bahawa min ukuran berfungsi seperti yang sepatutnya iaitu meningkat secara sekata merentasi skala pengukuran (-1.71, -0.70, -0.07, 0.62, 1.93)

Jadual 4.7

*Frekuensi kategori dan min ukuran bagi enam domain perkembangan*

Kategori (Skor)	Bilangan Pemerhatian (Frekuensi)	Min Ukuran	Outfit MNSQ
1	27	-1.71	1.3
2	30	-0.70	1.8
3	65	-0.07	1.1
4	181	0.62	1.0
5	561	1.93	0.9

Selain penggunaan frekuensi kategori dan peningkatan secara sekata dalam min ukuran untuk menentukan kesesuaian skala kategori yang digunakan, nilai kategori fit statistik (Lopez 1996; Wright & Master 1982) boleh juga digunakan. Nilai kesepadanan (fit statistik) memberikan satu lagi kriteria untuk menilai kualiti skala pengukuran. Menurut Linacre (1999), nilai outfit MNSQ yang melebihi 2 menunjukkan terdapat “*more misinformation than information*”, yang bermaksud bagi kategori tertentu terdapat banyak keadaan yang tidak sesuai dalam proses pengukuran. Kategori yang begini memerlukan penyiasatan lanjut secara empirikal dan merupakan calon yang sesuai untuk digabungkan dengan kategori yang berdekatan. Nilai outfit MNSQ dalam Jadual 4.7 menunjukkan nilai yang kurang daripada 2.0 bagi semua kategori bagi semua skor dalam instrumen. Ini menunjukkan bahawa semua kategori tersebut fit sepadan secara statistik dan memenuhi model PRPF.

Jadual 4.8 menunjukkan nilai kebolehpercayaan kanak-kanak bagi kajian ini adalah 0.95 dan indeks pengasingan adalah 4.35. Nilai ini adalah memenuhi kehendak model pengukuran Rasch. Linacre (2006) yang menyatakan kebolehpercayaan responden  $\geq 0.8$  dan nilai indeks pengasingan  $\geq 2$  adalah indeks yang diterima baik.

Jadual 4.8

*Nilai kebolehpercayaan bagi kanak-kanak dan nilai pengasingan*

Kanak-kanak	Ukuran	Nilai SE
11	2.39	0.16
13	2.39	0.16
2	2.19	0.14
3	1.42	0.10
1	1.24	0.09
12	1.02	0.09
Nilai kebolehpercayaan: 0.95		Nilai Pengasingan: 4.35

Jadual 4.9 menunjukkan statistik pengukuran faset penilai bagi domain perkembangan terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat di antara -0.95 logit (SE= 0.14) bagi penilai 13 iaitu penilai yang pemurah (*lenient*) sehingga 0.62 logit (SE=0.07) bagi penilai 18 iaitu penilai yang tegas. Hasil ujian khi kuasa dua menunjukkan terdapat perbezaan ketegasan yang signifikan ( $\chi^2=108.2$ ,  $df= 2.00$ ,  $p<0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Nilai infit dan outfit MNSQ dalam Jadual 4.9 menunjukkan setiap penilai dalam mengukur perkembangan kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Berdasarkan hasil dapatan, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.6-1.4. Ini menunjukkan bahawa setiap penilai menilai perkembangan kanak-kanak secara konsisten.

Jadual 4.9

*Nilai Ukuran Ketegasan, Perbezaan ketegasan dan Konsistensi Penilai*

Penilai	Ukuran	Infit MNSQ	Outfit MNSQ
18	0.62	1.19	1.01
8	0.33	0.97	1.02
13	-0.95	1.13	1.27
Nilai kebolehpercayaan: 0.98,		Nilai Pengasingan: 7.00	

Pemeriksaan terhadap analisis bias di antara kumpulan faset juga dilihat. Analisis bias dijalankan pada kajian kedua, sepertimana yang disarankan oleh Livingston (2006). Model PRPF digunakan untuk menyiasat interaksi bias antara faset. Istilah bias di antara penilai dan faset yang lain di sini bermaksud ketegasan atau kelonggaran penilai dalam membuat penilaian. Bias didefinisikan sebagai kecenderungan penilai dalam membuat penilaian yang rendah atau tinggi secara konsisten dari yang sepatutnya terhadap prestasi kanak-kanak (Engelhard, 1994). Wigglesworth (1993) menyatakan analisis bias dapat membantu penilai memahami proses penilaian yang berlaku dan dapat membuat penambahbaikan dalam pembentukan instrumen dan skala pengukuran. Dalam kajian ini, analisis bias yang dilakukan adalah dua interaksi bias antara kanak-kanak dan penilai/penilai, serta kanak-kanak dan item. Jadual 4.10 menunjukkan laporan bias kanak-kanak dan penilai. Mematuhi  $-1.96 \leq \text{nilai } t \leq +1.96$ . Merujuk bacaan probability  $\leq 0.05$  adalah signifikan bias.

Jadual 4.10  
*Laporan bias kanak-kanak dan penilai*

Kanak-kanak	Penilai	Nilai t	Nilai Prob.
2	13	1.63	0.1117
11	18	1.19	0.2421
12	18	1.16	0.2538
3	13	0.95	0.3464
13	18	0.89	0.3762
1	18	0.78	0.4403
12	18	0.68	0.4994
1	18	0.63	0.5329
2	18	0.49	0.6264
1	13	-0.02	0.9862
13	18	-0.04	0.9701
11	18	-0.05	0.9593
12	13	-0.75	0.4601
13	13	-0.91	0.3698
11	13	-1.20	0.2370
2	18	-1.27	0.2099
3	18	-1.78	0.0822
3	18	-2.46	0.0182



Hasil dapatan menunjukkan penilaian perkembangan kanak-kanak secara holistik boleh dilakukan dengan memberikan skor dan keputusan yang tepat berdasarkan pemerhatian penilai dan menunjukkan nilai indeks kesahan dan kebolehpercayaan. Hasil dapatan juga menunjukkan kanak-kanak yang berkebolehan tinggi dan rendah boleh dikenalpasti melalui nilai indeks tertinggi dan terendah dalam kumpulan umur yang berbeza dalam masa yang sama dinilai menggunakan instrumen yang sama. Kebolehan dan perkembangan kanak-kanak mempengaruhi kepelbagaian tahap prestasi kanak-kanak ke atas tugas (item). Kanak-kanak di dalam kumpulan umur yang sama dilihat mempunyai Zon Perkembangan Proksimal Vygotsky yang berbeza iaitu, tugas dilihat mudah bagi sesetengah kanak-kanak dan sedikit mencabar atau sukar atau mustahil dilakukan bagi sesetengah kanak-kanak yang lain (Ormrod, 2011). Kanak-kanak ini datang daripada latar belakang yang berbeza dari segi pengetahuan, kemahiran dan cara berfikir. Bergantung kepada pengetahuan tertentu dan tahap kebolehan, kanak-kanak yang berada dalam kumpulan umur yang berbeza memerlukan scaffolding atau sokongan untuk menjayakan sesuatu tugas (Lodewyk & Winne, 2005).

Hasil dapatan seterusnya menunjukkan apabila menggunakan guru sebagai penilai kepada kanak-kanak, aspek *severity dan leniency* dapat dikenal pasti menggunakan model Pengukuran Rasch Pelbagai Faset melalui nilai logit yang paling tinggi dan paling rendah. Persamaan prestasi penilai dilihat pada interaksi bias dan hasil kajian menunjukkan tiada interaksi bias yang signifikan menunjukkan persamaan ini meningkatkan kebolehpercayaan proses pemerhatian. Pendekatan yang digunakan untuk meningkatkan persamaan penilai adalah melalui 3 cara iaitu, latihan penilai, mentadbir skor kepada item prestasi yang sama oleh dua orang penilai atau lebih (Myford & Wolfe, 2002) dan menstabilkan kebolehpercayaan antara penilai (Kondo-

Brown, 2002). Oleh itu, penilai yang mempunyai perbezaan penilaian ini, memerlukan latihan untuk kesaksamaan penskoran terhadap kanak-kanak berpandukan rubrik dan standard prestasi.

Penilaian terhadap perkembangan kanak-kanak secara keseluruhan aspek bersifat subjektif kerana ia tidak menggunakan ujian berbentuk pensil dan kertas malah skor bergantung kepada pemerhati atau penilai. Oleh yang demikian, kajian ini menunjukkan alat penilaian atau instrumen yang tepat diperlukan agar data boleh dianalisa secara objektif dan serasi dengan model pengukuran tertentu yang boleh mentaksir prestasi perkembangan holistik kanak-kanak secara kuantitatif. Analisis data respon kanak-kanak berasaskan pentaksiran prestasi yang menggunakan model pengukuran Rasch Pelbagai Faset adalah berkesan untuk menilai perkembangan holistik kanak-kanak. Setiap item instrumen yang digunakan semasa menilai dapat dibuktikan dengan menguji serasi statistik. Selain itu, latihan dapat meningkatkan standard penilaian menggunakan pemerhatian dan persetujuan antara penilai. Kesimpulannya, Model Pengukuran Rasch Pelbagai Faset adalah sesuai digunakan dalam mengambil kira jumlah domain yang banyak dan diukur berasaskan pentaksiran prestasi iaitu kebolehan kanak-kanak, pengaruh penilai serta tahap kesukaran domain.

### **Rumusan**

Pentaksiran terhadap perkembangan kanak-kanak secara keseluruhan aspek bersifat subjektif kerana tidak menggunakan ujian berbentuk pensil dan kertas malah skor bergantung kepada pemerhati atau penilai. Oleh yang demikian, kajian ini menunjukkan alat penilaian atau instrumen yang tepat diperlukan agar data boleh dianalisis secara objektif dan serasi dengan model pengukuran tertentu yang boleh mentaksir prestasi perkembangan kanak-kanak secara kuantitatif. Analisis data respon kanak-kanak berasaskan pentaksiran prestasi yang menggunakan model PRPF adalah

berkesan untuk menilai perkembangan kanak-kanak. Setiap item instrumen yang digunakan semasa menilai dapat dibuktikan dengan menguji keserasian statistik. Selain itu, latihan dapat meningkatkan standard penilaian menggunakan pemerhatian dan persetujuan antara penilai. Kesimpulannya, Model pengukuran Rasch Pelbagai Faset adalah sesuai digunakan dalam mengambil kira jumlah domain yang banyak dan diukur berasaskan pentaksiran prestasi iaitu kebolehan kanak-kanak, ketegasan penilai serta tahap kesukaran domain.

Bahagian ini membentangkan dapatan hasil analisis yang dilakukan pada peringkat pengesahan yang terdiri daripada berdasarkan dapatan latihan guru. Dapatan ini membincangkan analisis dalam memberikan bukti-bukti pengesahan bagi memastikan instrumen berfungsi seperti yang sepatutnya. Analisis bias dilaksanakan pada peringkat ini bagi memastikan instrumen yang dihasilkan adalah sesuai untuk digunakan bagi mengukur dan membezakan perkembangan dan kebolehan kanak-kanak berdasarkan penilaian guru. Hasil kajian adalah merupakan satu instrumen yang telah diuji terhadap kesahan dan kebolehpercayaan. Selain itu menunjukkan kepentingan melakukan bengkel latihan guru untuk penskoran agar kesamarataan pandangan guru adalah sama semasa menilai prestasi kanak-kanak dan ia juga merupakan keperluan yang perlu dipenuhi untuk memenuhi jangkaan model PRPF ini.

## **Bab 5 Hasil Dapatan**

### **Pengenalan**

Bab ini akan menerangkan peringkat pengesahan di mana tujuan utamanya adalah memastikan instrumen yang dibina mempunyai kesahan dan kebolehpercayaan yang boleh diterima. Peringkat pengesahan bagi kajian ini dijalankan dalam dua fasa. Fokus fasa adalah lebih tertumpu kepada pembinaan instrumen pengukuran yang mempunyai ciri-ciri psikometrik yang bersesuaian untuk penggunaan skor prestasi perkembangan kanak-kanak dan merupakan peringkat pembinaan bagi pengesahan. Pada fasa kedua pula, penyelidik melihat semula sama ada instrumen yang dibina berfungsi sepertimana yang dijangkakan dan merupakan peringkat bagi pengesahan.

### **Responden Kajian**

Responden bagi kajian ini terdiri daripada 157 orang kanak-kanak taska dan 16 orang penilai. Daripada jumlah ini, seramai 80 orang kanak-kanak menggunakan Instrumen Perkembangan Holistik Kanak-kanak Taska 1 (IPHKT 1) dan 77 orang kanak-kanak menggunakan Instrumen Perkembangan Holistik Kanak-kanak Taska 2 (IPHKT 2). Dapatan kajian dalam bab ini diperolehi melalui guru yang dilatih untuk menilai setiap orang kanak-kanak.

### **Dapatan Kajian Pengesahan**

Bagi dapatan kajian pengesahan pusingan pertama, bukti-bukti pengesahan dikumpul berdasarkan kefungsi kategori skala dan laporan bagi pengukuran setiap faset. Bukti-bukti pengesahan ini adalah sepertimana yang dibincangkan di Bab 2. Analisis dibuat berdasarkan model Pengukuran Rasch Pelbagai Faset (PRPF) dan menggunakan perisian Facets 3.71.4.

**Kefungsian Kategori Skala.** Dalam menentukan kefungsian kategori skala yang digunakan bagi setiap konstruk dalam IPHKT 1 dan 2, tiga perkara yang dilihat. Pertama, bentuk taburan dengan melihat bilangan frekuensi setiap kategori; kedua, frekuensi bagi setiap kategori yang mestilah sekurang-kurangnya sepuluh; dan ketiga, nilai min ukuran yang semakin meningkat. Dapatan mengenai kefungsian kategori skala dibentangkan mengikut persoalan-persoalan kajian di bawah ini.

*Bagaimanakah kefungsian kelima-lima kategori skala yang digunakan bagi konstruk dalam IPHKT 1 dan 2?* Dalam kajian ini, lima kategori skala digunakan bagi kesemua konstruk dalam IPHKT 1 dan 2. Kanak-kanak yang menunjukkan respon tanpa bantuan bagi item pembelajaran yang dikaji dianggap mempunyai tahap pengukuran yang sangat tinggi dan ditandai dengan nilai skala pengukuran 5, manakala kanak-kanak yang tiada respon bagi item pembelajaran yang dikaji dianggap mempunyai kebolehan sangat rendah dan ditandai nilai skala pengukurannya.

Jadual 5.1 dan Jadual 5.2 menunjukkan bilangan pemerhatian dan min ukuran bagi skala pengukuran lima kategori bagi domain pembelajaran dalam IPHKT 1 dan 2. Daripada bilangan pemerhatian atau frekuensi bagi setiap kategori, dapat dilihat di sini bahawa ia adalah taburan normal tetapi agak pencong sedikit secara negatif, dan setiap kategori mempunyai lebih daripada sepuluh respons. Nilai min ukuran bagi kategori 1 adalah -1.82, bermakna min anggaran bagi kanak-kanak yang dinilai bagi kategori 1 merentasi item-item pembelajaran adalah -1.82 logit. Bagi kanak-kanak yang diberi nilai 2 bagi mana-mana item pembelajaran, min anggaran adalah -0.69. Dapat dilihat di sini bahawa min ukuran berfungsi seperti yang sepatutnya iaitu meningkat secara sekata merentasi skala pengukuran (-1.82, -0.69, -0.01, 0.68 dan 1.87).

Jadual 5.1  
*Frekuensi kategori dan min ukuran bagi IPHKT 1*

Kategori (Skor)	Bilangan Pemerhatian (Frekuensi)	Min Ukuran
1	2118	-1.82
2	1857	-0.69
3	3875	-0.01
4	7824	0.68
5	24901	1.87

Jadual 5.2  
*Frekuensi kategori dan min ukuran bagi IPHKT 2*

Kategori (Skor)	Bilangan Pemerhatian (Frekuensi)	Min Ukuran
1	534	-1.97
2	686	-0.90
3	3071	-0.15
4	10420	0.78
5	26304	2.40

Jadual 5.3 menunjukkan frekuensi kategori bagi tiga domain pembelajaran IPHKT 1 iaitu domain kognitif, domain bahasa dan domain fizikal. Jika dilihat secara kasar dari segi frekuensi kategori bagi ketiga-tiga domain pembelajaran, keadaan menunjukkan ketiga-tiga domain pembelajaran tersebut adalah bertabur secara normal tetapi menunjukkan sedikit pencong secara negatif. Setiap kategori juga menunjukkan bilangan frekuensi adalah melebihi 10. Dilihat daripada sudut min ukuran, keadaannya juga menunjukkan kesemua min ukuran meningkat secara sekata dan berfungsi sebagaimana yang sepatutnya.

Jadual 5.3  
*Frekuensi kategori dan min ukuran bagi domain kognitif, bahasa dan fizikal IPHKT 1*

Kategori (Skor)	Kognitif		Bahasa		Fizikal	
	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran
1	121	-2.48	290	-1.88	98	-2.11
2	263	-0.96	341	-0.63	176	-0.86
3	589	0.04	420	0.05	476	-0.04
4	1161	0.99	784	0.67	1374	0.82
5	3526	2.37	2762	1.77	5779	2.23

Jadual 5.4 menunjukkan frekuensi kategori bagi tiga lagi domain pembelajaran IPHKT 1 iaitu, domain kreativiti, domain sosioemosi dan domain rohani. Ketiga-tiga domain pembelajaran menunjukkan pola yang sama seperti domain sebelum ini iaitu mempunyai taburan yang sedikit pencong secara negatif, setiap kategori mempunyai frekuensi yang melebihi 10 dan min ukuran yang menunjukkan peningkatan yang sekata.

Jadual 5.4

*Frekuensi kategori dan min ukuran bagi domain kreativiti, sosioemosi dan rohani IPHKT 1*

Kategori (Skor)	Kreativiti		Sosioemosi		Rohani	
	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran
1	57	-2.00	807	-1.76	745	-1.83
2	109	-0.86	560	-0.69	402	-0.71
3	477	-0.07	1075	-0.04	811	-0.01
4	1048	0.80	2123	0.65	1257	0.70
5	2475	2.21	5163	1.89	3962	1.91

Jadual 5.5 menunjukkan frekuensi kategori bagi tiga domain pembelajaran IPHKT 2 iaitu domain kognitif, domain bahasa dan domain fizikal. Jika dilihat secara kasar dari segi frekuensi kategori bagi ketiga-tiga domain pembelajaran, keadaan menunjukkan ketiga-tiga domain pembelajaran tersebut adalah bertabur secara normal tetapi menunjukkan sedikit pencong secara negatif. Setiap kategori juga menunjukkan bilangan frekuensi adalah melebihi 10. Dilihat daripada sudut min ukuran, keadaannya juga menunjukkan kesemua min ukuran meningkat secara sekata dan berfungsi sebagaimana yang sepatutnya.

Jadual 5.5

*Frekuensi kategori dan min ukuran bagi domain kognitif, bahasa dan fizikal IPHKT 2*

Kategori (Skor)	Kognitif		Bahasa		Fizikal	
	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran	Frekuensi	Min Ukuran
1	25	-2.87	222	-1.72	22	-2.38
2	48	-1.57	128	-0.81	49	-1.16
3	427	-0.34	665	-0.17	362	-0.18
4	1953	1.49	1880	0.66	1255	1.08
5	5353	3.62	3799	2.22	4145	2.80

Jadual 5.6 menunjukkan frekuensi kategori bagi tiga lagi domain pembelajaran IPHKT 2 iaitu, domain kreativiti, domain sosioemosi dan domain rohani. Ketiga-tiga domain pembelajaran menunjukkan pola yang sama seperti domain sebelum ini iaitu mempunyai taburan yang sedikit pencong secara negatif, setiap kategori mempunyai frekuensi yang melebihi 10 dan min ukuran yang menunjukkan peningkatan yang sekata.

Jadual 5.6  
*Frekuensi kategori dan min ukuran bagi domain kreativiti, sosioemosi dan rohani IPHKT 2*

Kategori (Skor)	Kreativiti		Sosioemosi		Rohani	
	Frek	Min Ukuran	Frek	Min Ukuran	Frek	Min Ukuran
1	81	-2.61	89	-1.99	85	-1.97
2	245	-1.04	78	-1.01	113	-.89
3	660	0.2	528	-.27	380	-.15
4	1749	1.05	2065	.84	1332	.77
5	4393	2.58	4012	2.71	3184	2.43

*Adakah setiap kategori sepadan (fit) dengan model pengukuran Rasch pelbagai faset (PRPF)?* Selain daripada menggunakan frekuensi kategori dan peningkatan secara sekata dalam min ukuran untuk menentukan kesesuaian penggunaan skala kategori, nilai kesepadanan (fit) statistik juga digunakan (Lopez, 1996, Wright & Masters, 1982). Nilai kesepadanan (fit) statistik memberikan satu lagi kriteria untuk menilai kualiti skala pengukuran. Kategori yang begini memerlukan penyiasatan lanjut secara empirikal dan mungkin merupakan calon yang sesuai untuk digabungkan dengan kategori yang berdekatan. Jadual 5.7 dan 5.8 menunjukkan nilai outfit MNSQ bagi setiap kategori bagi semua enam konstruk pembelajaran dalam IPHKT 1 dan 2.



Jadual 5.7

*Nilai Outfit MNSQ bagi enam konstruk pembelajaran kanak-kanak dalam IPHKT 1*

Domain	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
Kognitif	3.9	0.8	0.8	0.7	1.0
Bahasa	2.7	0.5	0.9	0.9	0.9
Fizikal	2.3	1.6	1.0	0.8	1.0
Kreativiti	2.7	1.9	0.7	0.7	1.0
Sosioemosi	2.2	0.9	0.6	0.7	0.9
Rohani	2.4	0.7	0.9	0.8	0.9

Jadual 5.8

*Nilai Outfit MNSQ bagi enam konstruk pembelajaran kanak-kanak dalam IPHKT 2*

Domain	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
Kognitif	2.9	2.3	1.4	0.7	1.0
Bahasa	1.5	1.1	0.7	1.0	1.0
Fizikal	2.8	2.6	1.2	0.6	1.0
Kreativiti	4.1	1.0	0.8	0.7	1.0
Sosioemosi	2.2	1.7	0.9	1.0	1.0
Rohani	2.8	0.9	0.9	0.8	1.0

Nilai outfit MNSQ yang ditunjukkan pada Jadual 5.7 dan 5.8 di atas menunjukkan semua domain dalam IPHKT 1 dan IPHKT 2 menunjukkan bacaan nilai kategori 1 yang lebih daripada 2.0 kecuali domain bahasa IPHKT 2. Manakala bagi semua kategori bagi semua konstruk domain pembelajaran dalam IPHKT 1 dan 2 bagi kategori 2 kecuali domain kognitif dan Fizikal IPHKT 2 menunjukkan bacaan nilai kurang daripada 2.0. Ini menunjukkan kategori tersebut fit secara statistik dan memenuhi model pengukuran PRPF.

***Adakah nilai ambang (threshold) menunjukkan corak hierarki bagi skala pengukuran yang digunakan bagi setiap kategori (skala rating) dalam IPHKT ?***

Nilai ambang (*threshold*) atau *step calibrations* adalah satu lagi kriteria penting yang perlu dilihat dalam menentukan kesesuaian penggunaan kategori bagi skala pengukuran yang digunakan. Nilai *step calibrations* adalah anggaran kesukaran bagi memilih kategori respons berbanding kategori bersebelahan. Sebagai contoh, bagaimanakah kesukaran memilih kategori “sangat berkebolehan” berbanding

“berkebolehan”. Seperti nilai min ukuran, nilai *step calibrations* juga perlu meningkat secara sekata. Nilai yang tidak meningkat secara sekata merentasi skala pengukuran adalah dianggap tidak teratur.

Magnitud bagi jarak antara nilai anggaran ambang (*threshold*) juga adalah penting. Jarak ambang adalah untuk menunjukkan setiap langkah mendefinisikan posisi yang berbeza bagi kategori. Ini menunjukkan bahawa nilai anggaran adalah terlalu dekat atau terlalu jauh bagi skala logit yang digunakan. Menurut Linacre (1999), nilai ambang perlu meningkat sekurang-kurangnya sebanyak 1.4 logit (bagi skala pengukuran 5 kategori) untuk menunjukkan perbezaan jarak antara kategori yang digunakan dan tidak melebihi 5 logit agar dapat mengelakkan jurang yang sangat besar di antara kategori yang digunakan.

Kaedah secara visual juga boleh digunakan untuk menyiasat perbezaan antara nilai ambang dengan melihat lengkung kebarangkalian yang menunjukkan kebarangkalian bagi satu kategori skala pengukuran. Setiap kategori seharusnya mempunyai puncak yang berbeza dalam lengkung kebarangkalian bagi menunjukkan setiap kategori respons yang paling mungkin bagi sebahagian pembolehubah yang diukur (Bond & Fox 2001). Kategori yang dilihat “mendatar” pada graf adalah kategori yang tidak mungkin digunakan dan ini menunjukkan kategori tersebut tidak membantu dalam mendefinisikan perbezaan titik pada kategori (skor). Oleh itu, nilai ambang yang terlalu rapat atau tidak teratur dapat dilihat secara visual dengan bentuk lengkung kebarangkalian yang mendatar dianggap sebagai bermasalah.

Jadual 5.9 dan 5.10 adalah nilai ambang bagi 5 kategori skala pengukuran domain pembelajaran IPHKT 1 dan 2. Nilai ambang bagi 5 kategori ini menunjukkan peningkatan yang kurang sekata khususnya bagi kategori 3 dan kategori 4 yang menunjukkan julat kurang daripada 1.0. Ini menunjukkan bahawa kategori ini masih

memenuhi model pengukuran Rasch namun perlu dibaiki kesesuaian penggunaan dua kategori ini sama ada dikekalkan asing atau digabungkan.

Jadual 5.9

*Nilai ambang (threshold) bagi 5 kategori skala pengukuran bagi kesemua domain dalam IPHKT 1*

Domain	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
Kognitif	Tiada	-1.76	-0.42	0.49	1.72
Bahasa	Tiada	-1.24	-0.25	0.34	1.21
Fizikal	Tiada	-1.49	-0.42	0.35	1.55
Kreativiti	Tiada	-1.45	-0.44	0.32	1.54
Sosioemosi	Tiada	-1.23	-0.34	0.27	1.28
Rohani	Tiada	-1.28	-0.34	0.31	1.30

Jadual 5.10

*Nilai ambang (threshold) bagi 5 kategori skala pengukuran bagi kesemua domain dalam IPHKT 2*

Domain	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
Kognitif	Tiada	-2.27	-0.99	0.49	2.71
Bahasa	Tiada	-1.28	-0.48	0.18	1.47
Fizikal	Tiada	-1.80	-0.67	0.39	2.02
Kreativiti	Tiada	-1.87	-0.47	0.50	1.86
Sosioemosi	Tiada	-1.52	-0.64	0.18	1.86
Rohani	Tiada	-1.45	-0.51	0.24	1.65

**Laporan Pengukuran Domain.** Dapatan mengenai laporan pengukuran domain adalah seperti di bawah khususnya berkenaan dengan ciri unidimensionaliti.

Rujuk lampiran

*Adakah setiap domain IPHKT 1 dan 2 mempunyai ciri-ciri unidimensionaliti?*

Jadual 5.11

*Nilai pematuhan Unidimensionaliti Berdasarkan Rasch FACETS*

Domain	IPHKT 1	IPHKT 2
	>20%	>20%
Kognitif	49.34	41.58
Bahasa	53.15	40.94
Fizikal	42.21	32.72
Kreativiti	30.78	40.67
Sosioemosi	51.83	38.94
Rohani	63.54	40.06

**Laporan Pengukuran Kanak-kanak.** Dapatan mengenai laporan pengukuran pelajar dibentangkan mengikut susunan persoalan-persoalan kajian seperti di bawah ini.

*Apakah nilai kebolehpercayaan kanak-kanak bagi IPHKT 1 dan 2?* Jadual 5.12 menunjukkan nilai kebolehpercayaan kanak-kanak bagi IPHKT 1 kajian ini adalah di antara 0.89 hingga 0.97 dengan indeks pengasingan di antara 2.90 hingga 5.43. Manakala nilai kebolehpercayaan kanak-kanak bagi IPHKT 2 kajian ini adalah di antara 0.90 hingga 0.96 dengan indeks pengasingan di antara 3.04 hingga 4.69. Nilai ini adalah memenuhi kehendak model pengukuran Rasch. Linacre (2006) yang menyatakan kebolehpercayaan responden  $\geq 0.8$  dan nilai indeks pengasingan  $\geq 2$  adalah indeks yang diterima baik.

Jadual 5.12

*Nilai kebolehpercayaan bagi kanak-kanak dan nilai indeks pengasingan IPHKT 1*

Domain	Nilai Kebolehpercayaan	Indeks Pengasingan
Kognitif	.93	3.75
Bahasa	.92	3.51
Fizikal	.93	3.77
Kreativiti	.89	2.90
Sosioemosi	.97	5.43
Rohani	.95	4.34

Jadual 5.13

*Nilai kebolehpercayaan bagi kanak-kanak dan nilai indeks pengasingan IPHKT 2*

Domain	Nilai Kebolehpercayaan	Indeks Pengasingan
Kognitif	.96	4.69
Bahasa	.92	3.49
Fizikal	.90	3.04
Kreativiti	.93	3.64
Sosioemosi	.96	4.69
Rohani	.92	3.30

*Apakah ukuran yang dilaksanakan dengan menggunakan IPHKT 1 dan 2 dapat mengasingkan kanak-kanak kepada keupayaan yang berbeza?* Jadual 5.14 menunjukkan laporan ukuran kanak-kanak bagi enam domain yang terdapat dalam IPHKT iaitu julat ukuran kanak-kanak, nilai RMSE, nilai indeks pengasingan dan nilai

khikuasa dua. Bagi domain kognitif julat ukuran kanak-kanak adalah di antara -0.42 (ralat piawai= 0.06, kanak-kanak 205) iaitu ukuran kanak-kanak yang terendah dan 3.93 (ralat piawai = 0.50, kanak-kanak 340). Nilai RMSE bagi pengukuran kanak-kanak domain kognitif adalah 0.26. Nilai indeks pengasingan adalah 3.75 menunjukkan kanak-kanak ini dapat dipisah atau diasingkan kepada keupayaan yang berbeza. Ini disahkan lagi dengan nilai khikuasdua yang signifikan di mana bagi domain kognitif nilai  $\chi^2 = 2461.7$ ,  $p < 0.01$ ,  $df = 79$ .

Lima domain yang lain juga menunjukkan dapatan yang tidak begitu berbeza. Nilai RMSE bagi setiap domain yang lain adalah berada di antara 0.14 (domain sosioemosi) dan 0.29 (domain fizikal). Nilai indeks pengasingan dan nilai khikuasa dua yang signifikan juga menunjukkan bahawa ukuran kanak-kanak yang diperolehi dapat memisahkan kanak-kanak kepada keupayaan yang berbeza. Nilai indeks pengasingan yang tertinggi adalah bagi domain sosioemosi iaitu 5.43 dan nilai  $\chi^2 = 3297.3$  ( $p < 0.01$ ,  $df = 79$ ), dan nilai indeks pengasingan terendah adalah bagi domain kreativiti iaitu 2.90 dan nilai  $\chi^2 = 1018.4$  ( $p < 0.01$ ,  $df = 79$ ).

Manakala Jadual 5.15 menunjukkan nilai indeks pengasingan yang tertinggi bagi IPHKT 2 adalah 4.69 bagi domain kognitif ( $\chi^2 = 2364.9$ ,  $p < 0.01$ ,  $df = 76$ ) dan sosioemosi ( $\chi^2 = 2364.9$ ,  $p < 0.01$ ,  $df = 76$ ), dan nilai indeks pengasingan terendah adalah bagi domain fizikal iaitu 3.04 dan nilai  $\chi^2 = 1143.6$  ( $p < 0.01$ ,  $df = 76$ ). Ini menunjukkan bahawa item-item bagi setiap domain dalam IPHKT 1 dan 2 telah dapat memisah atau mengasingkan kanak-kanak kepada beberapa tahap keupayaan yang berbeza.

Jadual 5.14

Laporan pengukuran kanak-kanak bagi enam domain pembelajaran ( $n=80$ ) dengan menggunakan IPHKT 1

Domain	Julat logit	Ralat piawai (SE)	Nilai RMSE	Nilai indeks pengasingan	Nilai dengan $p<0.01$ $df=79$	$\chi^2$
Kognitif	-0.42 – 3.93	0.24	0.26	3.75	2461.7	
Bahasa	-0.31 – 3.13	0.21	0.23	3.51	992.2	
Fizikal	-0.40 – 4.76	0.33	0.29	3.77	2715.6	
Kreativiti	-0.06 – 3.10	0.60	0.24	2.90	1018.4	
Sosioemosi	-0.48 – 2.49	0.13	0.14	5.43	3297.3	
Rohani	-0.36 – 2.42	0.15	0.16	4.34	1888.4	

Jadual 5.15

Laporan pengukuran kanak-kanak bagi enam domain pembelajaran ( $n=77$ ) dengan menggunakan IPHKT 2

Domain	Julat ukuran	Ralat piawai (SE)	Nilai RMSE	Nilai indeks pengasingan	Nilai dengan $p<0.01$ $df=76$	$\chi^2$
Kognitif	-0.57 – 7.12	0.28	0.31	4.69	2364.9	
Bahasa	-0.51 – 4.90	0.32	0.31	3.49	2324.8	
Fizikal	0.47 – 5.32	0.36	0.33	3.04	1143.6	
Kreativiti	0.07 – 4.65	0.26	0.29	3.64	2360.3	
Sosioemosi	-0.78 – 5.24	0.31	0.23	4.69	2202.1	
Rohani	0.28 – 5.07	0.34	0.29	3.30	1125.8	

Bagi melihat perbezaan keupayaan kanak-kanak mengikut umur, perbezaan *mean measure* umur 2 dan 3 tahun bagi IPHKT 1 dicari. Bagi IPHKT 2, *mean measure* umur 3 dan 4 tahun dicari. Paparan perbezaan keupayaan bagi peringkat umur ini dilihat di dalam bentuk graf *boxplot*. *Mean measure* bagi kumpulan umur 2 tahun adalah sebanyak 1.162 manakala kumpulan *mean measure* umur 3 tahun adalah sebanyak 1.438. Dapatan menunjukkan kumpulan kanak-kanak 3 tahun mempunyai kebolehan yang lebih tinggi berbanding kumpulan umur 2 tahun. Manakala *mean measure* bagi kumpulan umur 3 tahun adalah sebanyak 2.592 manakala kumpulan *mean measure* umur 4 tahun adalah sebanyak 1.823. Dapatan menunjukkan kebolehan kanak-kanak 3 tahun lebih tinggi berbanding kumpulan umur 4 tahun. Kedudukan perbezaan kebolehan ini boleh dilihat dalam paparan graf *boxplot* di Lampiran G, H.

**Laporan Pengukuran Penilai.** Dapatan mengenai laporan pengukuran penilai dibentangkan mengikut susunan persoalan-persoalan kajian seperti di bawah ini.

***Apakah nilai peratus persetujuan antara penilai?*** Nilai peratus persetujuan penilai dilihat dari segi nilai peratus persetujuan antara penilai yang menunjukkan penilai menilai kanak-kanak berdasarkan konstruk dengan betul. Jadual 5.16 dan 5.17 menunjukkan nilai peratus persetujuan bagi enam domain IPHKT 1 dan 2.

***Nilai kebolehpercayaan penilai.*** Nilai peratus persetujuan antara penilai yang diperolehi melalui analisis Rasch dengan menggunakan perisian FASETS telah digunakan juga untuk menentukan nilai kebolehpercayaan. Nilai ini hampir serupa dengan Cohen's Kappa sebagai indikator kepada kebolehpercayaan antara penilai. Rumus pengiraan nilai kebolehpercayaan bagi kajian ini adalah seperti yang terdapat dalam bab tiga. Nilai kebolehpercayaan bagi kajian ini adalah seperti yang terdapat di Jadual 5.16 menunjukkan nilai kebolehpercayaan dalam IPHKT 1 antara penilai yang tinggi iaitu di antara 0.92 bagi pengukuran yang dibuat oleh penilai terhadap domain kognitif kanak-kanak dan 0.98 iaitu bagi domain Fizikal dan Rohani. Nilai kebolehpercayaan dalam Jadual 5.17 menunjukkan nilai kebolehpercayaan antara penilai IPHKT 2 yang tinggi iaitu di antara 0.95 bagi pengukuran yang dibuat oleh penilai terhadap domain Rohani dan 0.99 iaitu bagi domain kognitif dan sosioemosi.

***Adakah guru penilai bertindak sebagai penilai yang bebas?*** Bagi menjawab persoalan kajian ini, sekali lagi nilai peratus persetujuan sebenar (*observed*) dan nilai peratus persetujuan yang dijangkakan (*expected*) dilihat bagi menentukan bagaimana penilai bertindak dalam membuat penilaian. Jadual 5.16 dan 5.17 menunjukkan peratus persetujuan antara penilai bagi setiap domain yang diperhatikan, diikuti oleh peratus jangkakan menurut model pengukuran Rasch dan juga nilai kesepadanan berdasarkan model pengukuran Rasch. Jadual 5.16 menunjukkan peratus persetujuan

yang paling tinggi di antara penilai bagi IPHKT 1 adalah domain Fizikal iaitu 72.2% dan yang paling rendah adalah bagi mengukur domain sosioemosi iaitu 48.2%.

Jadual 5.17 pula menunjukkan peratus persetujuan yang paling tinggi antara penilai dalam IPHKT 2 adalah mengukur domain fizikal iaitu 63.7 % dan persetujuan yang paling rendah adalah mengukur domain bahasa iaitu 46.6% . Dalam Jadual 5.16 dan Jadual 5.17 juga dapat diperhatikan bahawa nilai sebenar (*observed*) adalah melebihi nilai jangkakan (*expected*) bagi setiap domain yang dinilai oleh penilai dan ini menunjukkan bahawa penilai telah bertindak seperti penilai yang terlatih, namun begitu mereka tidak bergantung kepada rakan penilai yang lain dalam membuat penilaian (peratus persetujuan antara penilai berdasarkan nilai peratus sebenar (*observed*) dan jangkakan (*expected*)). Ini memenuhi kehendak penyelidik dalam memastikan guru penilai membuat penilaian yang adil di mana setiap skor yang diberi adalah atas kehendak sendiri dan tidak dipengaruhi oleh penilai yang lain.

Jadual 5.16

*Nilai peratus persetujuan di antara penilai, nilai fit dan kebolehppercayaan IPHKT 1*

Domain	Nilai persetujuan (%)	Nilai jangkakan (%)	Fit	Nilai Kebolehppercayaan
Kognitif	60.1	51.9	1.07	0.92
Bahasa	58.3	51.3	1.05	0.96
Fizikal	72.2	68.9	1.10	0.98
Kreativiti	53.5	48.1	1.05	0.95
Sosioemosi	48.2	38.6	1.07	0.96
Rohani	63.7	48.9	1.07	0.98

Jadual 5.17

*Nilai peratus persetujuan di antara penilai, nilai fit dan kebolehppercayaan IPHKT 2*

Domain	Nilai persetujuan (%)	Nilai jangkakan (%)	Fit	Nilai Kebolehppercayaan
Kognitif	59.3	58.4	1.00	0.99
Bahasa	46.6	41.6	1.05	0.96
Fizikal	63.7	60.3	1.16	0.97
Kreativiti	50.0	46.0	1.11	0.97
Sosioemosi	59.1	52.9	1.07	0.99
Rohani	55.1	49.0	1.08	0.95



Bagi menjawab tiga persoalan kajian yang seterusnya, iaitu ketegasan penilai dan sama ada penilai-penilai yang menilai kanak-kanak mempunyai ketegasan yang berbeza mengenai penilaian yang konsisten, penyelidik akan menjawab ketiga-tiga soalan kajian secara serentak berdasarkan kepada setiap domain dalam IPHKT 1 dan 2. Tiga persoalan kajian tersebut adalah:

***Apakah ukuran ketegasan penilai setiap domain dalam IPHKT 1 dan 2?***

***Apakah terdapat perbezaan ketegasan yang signifikan antara penilai dalam menilai kanak-kanak? Adakah setiap penilai menilai dengan konsisten bagi setiap domain dalam IPHKT 1 dan 2?***

*Domain Kognitif 1.* Bagi ukuran ketegasan Jadual 5.18 menunjukkan statistik pengukuran faset penilai bagi domain kognitif IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.52 logit (SE= 0.09) bagi penilai 19 iaitu penilai yang pemurah (*lenient*) sehingga 0.55 logit (SE= 0.10) bagi penilai 27 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.12 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain kognitif IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.06 sehingga 0.10.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 168.8$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Bagi konsistensi penilai nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain kognitif, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat seorang penilai iaitu penilai 19 yang mempunyai nilai infit MNSQ yang

melebihi 1.5 (infit MNSQ = 1.57). Lima belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain kognitif 1 setiap kanak-kanak secara konsisten.

Jadual 5.18

*Laporan ukuran ketegasan penilai bagi domain kognitif IPHKT 1 mengikut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 27	0.55	0.10	0.83	0.80
PENILAI 26	0.52	0.10	0.86	0.84
PENILAI 8	0.34	0.07	1.09	1.23
PENILAI 9	0.16	0.08	1.21	1.32
PENILAI 4	0.05	0.07	0.83	0.64
PENILAI 3	0.05	0.07	0.84	0.72
PENILAI 30	0.03	0.07	0.96	0.80
PENILAI 31	0.02	0.06	1.00	0.97
PENILAI 29	0.01	0.07	0.86	0.76
PENILAI 25	-0.09	0.09	1.26	1.13
PENILAI 32	-0.13	0.06	1.13	1.17
PENILAI 13	-0.13	0.09	1.20	0.96
PENILAI 24	-0.14	0.09	0.75	0.76
PENILAI 23	-0.23	0.08	1.52	1.38
PENILAI 22	-0.50	0.10	1.23	1.16
PENILAI 19	-0.52	0.09	1.57**	1.49

*Nota.* Nilai khi-kuasadua: 168.8 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Kognitif 2.* Bagi ukuran ketegasan, Jadual 5.19 menunjukkan statistik pengukuran faset penilai bagi domain kognitif IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -1.37 logit (SE= 0.14) bagi penilai 20 iaitu penilai yang pemurah (*lenient*) sehingga 1.89 logit (SE= 0.10) bagi penilai 26 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 3.26 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain kognitif IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.07 sehingga 0.14.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 168.8$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Bagi Konsistensi penilai, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain kognitif, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat tiga orang penilai iaitu penilai 32 (infit MNSQ = 1.57, outfit MNSQ = 1.72), penilai 19 (infit MNSQ = 1.52) dan penilai 20 (infit MNSQ = 1.55) yang mempunyai nilai infit MNSQ yang melebihi 1.5. Tiga belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain kognitif setiap kanak-kanak secara konsisten.

Jadual 5.19

*Laporan ukuran ketegasan penilai bagi domain kognitif IPHKT 2 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 26	1.89	0.10	0.62	0.60
PENILAI 31	1.72	0.07	1.21	0.97
PENILAI 27	1.47	0.10	0.54	0.59
PENILAI 29	0.85	0.09	1.13	0.89
PENILAI 32	0.43	0.09	1.57**	1.72**
PENILAI 30	0.03	0.11	1.24	1.17
PENILAI 1	0.02	0.09	0.96	0.97
PENILAI 24	-0.39	0.10	0.84	0.78
PENILAI 8	-0.42	0.09	1.18	0.97
PENILAI 15	-0.46	0.09	0.82	0.61
PENILAI 3	-0.54	0.08	1.15	0.92
PENILAI 11	-0.56	0.10	0.89	0.70
PENILAI 9	-0.63	0.10	0.93	0.87
PENILAI 4	-0.96	0.09	1.37	0.83
PENILAI 19	-1.08	0.13	1.52**	1.05
PENILAI 20	-1.37	0.14	1.55**	1.03

*Nota.* Nilai khi-kuasdua: 1665.6 ;  $df$ : 15; Signifikan (kebarangkalian) : 0.00

*Domain Bahasa 1.* Bagi ukuran ketegasan, Jadual 5.20 menunjukkan statistik pengukuran faset penilai bagi domain bahasa IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.59 logit (SE= 0.09) bagi penilai 22 iaitu penilai yang pemurah (*lenient*) sehingga 0.62 logit (SE= 0.08) bagi penilai 27 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.21 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain bahasa IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.05 sehingga 0.09.

Bagi perbezaan ketegasan, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 367.1$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain bahasa, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat seorang penilai iaitu penilai 19 (infit MNSQ = 1.52, outfit MNSQ = 2.17) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Lima belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain bahasa setiap kanak-kanak secara konsisten.

Jadual 5.20

Laporan ukuran ketegasan penilai bagi domain bahasa IPHKT 1 menurut urutan ukuran ketegasan

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 27	0.62	0.08	1.02	1.07
PENILAI 26	0.60	0.08	0.76	0.81
PENILAI 24	0.37	0.07	0.92	0.84
PENILAI 31	0.33	0.06	0.84	0.84
PENILAI 32	0.27	0.06	0.82	0.91
PENILAI 25	0.12	0.08	0.88	1.14
PENILAI 8	0.08	0.07	1.16	1.53
PENILAI 30	0.05	0.05	0.78	0.84
PENILAI 29	0.03	0.06	0.72	0.72
PENILAI 9	-0.02	0.07	1.36	1.72
PENILAI 19	-0.19	0.08	1.92**	2.17**
PENILAI 3	-0.29	0.06	0.91	1.06
PENILAI 23	-0.37	0.08	1.45	1.24
PENILAI 4	-0.47	0.07	0.74	0.76
PENILAI 13	-0.52	0.09	1.11	0.91
PENILAI 22	-0.59	0.09	1.38	1.28

Nota. Nilai khi-kuasadua: 367.1 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Bahasa 2.* Bagi ukuran ketegasan, Jadual 5.21 menunjukkan statistik pengukuran faset penilai bagi domain bahasa IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.43 logit (SE= 0.09) bagi penilai 19 iaitu penilai yang pemurah (*lenient*) sehingga 0.72 logit (SE= 0.07) bagi penilai 26 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.15 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain bahasa IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.06 sehingga 0.09.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 383.6$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Bagi konsistensi penilai, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain bahasa, nilai infit dan outfit MNSQ

bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat seorang penilai iaitu penilai 20 (infit MNSQ = 2.02, outfit MNSQ = 1.69) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5 manakala penilai 27 (infit MNSQ = 0.49) mempunyai nilai infit dan outfit MNSQ yang kurang 0.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain bahasa setiap kanak-kanak secara konsisten.

Jadual 5.21

*Laporan ukuran ketegasan penilai bagi domain bahasa IPHKT 2 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 26	0.72	0.07	0.57	0.58
PENILAI 31	0.64	0.06	0.76	0.64
PENILAI 27	0.43	0.08	0.49**	0.67
PENILAI 29	0.20	0.07	0.94	0.94
PENILAI 1	0.07	0.07	0.96	0.85
PENILAI 20	0.06	0.07	2.02**	1.69**
PENILAI 3	-0.05	0.06	0.93	0.99
PENILAI 8	-0.13	0.06	0.76	0.77
PENILAI 24	-0.15	0.07	1.00	0.92
PENILAI 32	-0.15	0.07	1.10	1.48
PENILAI 9	-0.15	0.06	0.81	0.80
PENILAI 4	-0.18	0.06	1.31	0.99
PENILAI 11	-0.21	0.07	1.27	1.00
PENILAI 15	-0.32	0.07	1.17	1.03
PENILAI 30	-0.35	0.08	1.06	1.17
PENILAI 19	-0.43	0.09	1.71**	1.49

*Nota.* Nilai khi-kuasadua: 383.6 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Fizikal 1.* Bagi ukuran ketegasan, Jadual 5.22 menunjukkan statistik pengukuran faset penilai bagi domain fizikal IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.93 logit (SE= 0.10) bagi penilai 24 iaitu penilai yang pemurah (*lenient*) sehingga 1.06 logit (SE= 0.10) bagi penilai 27 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.99 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat

piawai bagi penilaian penilai terhadap domain fizikal IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.05 sehingga 0.10.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 681.5$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Bagi konsistensi penilai, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain fizikal, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat dua orang penilai iaitu penilai 22 (infit MNSQ = 1.56, outfit MNSQ = 1.71) dan penilai 23 (infit MNSQ = 1.87) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain fizikal setiap kanak-kanak secara konsisten.

Jadual 5.22

*Laporan ukuran ketegasan penilai bagi domain fizikal IPHKT 1 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 27	1.06	0.10	0.81	0.79
PENILAI 26	1.05	0.10	0.68	0.69
PENILAI 31	0.63	0.05	0.90	0.88
PENILAI 32	0.38	0.06	0.89	1.13
PENILAI 30	0.21	0.08	0.90	0.67
PENILAI 29	0.17	0.08	0.96	0.72
PENILAI 22	0.08	0.09	1.56**	1.71**
PENILAI 13	-0.06	0.10	1.16	0.92
PENILAI 8	-0.08	0.07	1.18	1.08
PENILAI 3	-0.16	0.06	1.05	0.90
PENILAI 9	-0.19	0.07	1.20	0.95
PENILAI 25	-0.48	0.08	0.88	0.86
PENILAI 4	-0.53	0.07	1.32	0.97
PENILAI 23	-0.56	0.08	1.87**	1.36
PENILAI 19	-0.59	0.09	1.45	1.27
PENILAI 24	-0.93	0.10	0.84	1.65

*Nota.* Nilai khi-kuasadua: 681.5 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Fizikal 2.* Bagi ukuran ketegasan, Jadual 5.23 menunjukkan statistik pengukuran faset penilai bagi domain fizikal IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.96 logit (SE= 0.14) bagi penilai 20 iaitu penilai yang pemurah (*lenient*) sehingga 1.17 logit (SE= 0.11) bagi penilai 27 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 2.13 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain fizikal IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.08 sehingga 0.14.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 570.5$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain fizikal, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat seorang penilai iaitu penilai 20 (infit MNSQ = 1.96) yang mempunyai nilai infit MNSQ yang melebihi 1.5. Lima belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain fizikal setiap kanak-kanak secara konsisten.



Jadual 5.23

Laporan ukuran ketegasan penilai bagi domain fizikal IPHKT 2 menurut urutan ukuran ketegasan

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 27	1.17	0.11	1.01	0.97
PENILAI 26	1.17	0.11	1.05	1.05
PENILAI 31	0.94	0.09	0.84	0.74
PENILAI 29	0.47	0.10	1.22	1.42
PENILAI 32	0.45	0.10	1.00	1.01
PENILAI 30	-0.03	0.12	1.21	1.76
PENILAI 3	-0.15	0.08	0.98	0.90
PENILAI 1	-0.20	0.10	1.29	0.95
PENILAI 15	-0.21	0.09	1.30	0.90
PENILAI 8	-0.22	0.10	0.75	0.61
PENILAI 11	-0.27	0.10	1.14	0.71
PENILAI 4	-0.40	0.08	1.04	0.76
PENILAI 24	-0.40	0.10	1.24	0.86
PENILAI 9	-0.43	0.10	1.10	0.81
PENILAI 19	-0.93	0.14	1.41	0.99
PENILAI 20	-0.96	0.14	1.96**	1.01

Nota. Nilai khi-kuasadua: 570.5 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Kreativiti 1.* Bagi ukuran ketegasan, Jadual 5.24 menunjukkan statistik pengukuran faset penilai bagi domain kreativiti IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.71 logit (SE= 0.10) bagi penilai 19 iaitu penilai yang pemurah (*lenient*) sehingga 0.59 logit (SE= 0.06) bagi penilai 32 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.30 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain kreativiti IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.06 sehingga 0.11.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 332.2$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain kreativiti, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan

1.5. Terdapat dua orang penilai iaitu penilai 19 (infit MNSQ =1.90, outfit MNSQ = 1.69) dan penilai 30 (infit MNSQ =1.72, outfit MNSQ = 1.69) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain kreativiti setiap kanak-kanak secara konsisten.

Jadual 5.24

*Laporan ukuran ketegasan penilai bagi domain kreativiti IPHKT 1 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 32	0.59	0.06	0.99	1.01
PENILAI 26	0.59	0.10	0.53	0.56
PENILAI 27	0.47	0.11	0.71	0.72
PENILAI 31	0.44	0.07	0.76	0.77
PENILAI 8	0.31	0.08	0.68	0.68
PENILAI 9	0.18	0.08	1.10	1.10
PENILAI 30	0.10	0.08	1.72**	1.69**
PENILAI 25	0.02	0.09	0.86	0.93
PENILAI 24	-0.11	0.09	0.58	0.58
PENILAI 29	-0.15	0.09	1.50	1.30
PENILAI 3	-0.16	0.08	0.87	0.97
PENILAI 4	-0.20	0.08	1.30	1.30
PENILAI 23	-0.27	0.09	1.41	1.23
PENILAI 13	-0.52	0.09	0.82	0.94
PENILAI 22	-0.56	0.10	1.15	1.27
PENILAI 19	-0.71	0.10	1.90**	1.69**

*Nota.* Nilai khi-kuasdua: 332.2 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Kreativiti.* Bagi ukuran ketegasan, Jadual 5.25 menunjukkan statistik pengukuran faset penilai bagi domain kreativiti IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.64 logit (SE= 0.09) bagi penilai 19 iaitu penilai yang pemurah (*lenient*) sehingga 0.77 logit (SE= 0.08) bagi penilai 27 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.41 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain kreativiti IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.06 sehingga 0.09.

Perbezaan ketegasan pula, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 545.7$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain kreativiti, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat dua orang penilai iaitu penilai 20 (infit MNSQ =2.34, outfit MNSQ = 1.81) dan penilai 19 (infit MNSQ =1.56) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain kreativiti setiap kanak-kanak secara konsisten.

#### Jadual 5.25

*Laporan ukuran ketegasan penilai bagi domain kreativiti IPHKT 2 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 27	0.77	0.08	0.54	0.63
PENILAI 31	0.75	0.06	0.92	0.80
PENILAI 26	0.73	0.08	0.66	0.65
PENILAI 32	0.27	0.07	0.86	1.27
PENILAI 29	0.22	0.07	1.25	0.98
PENILAI 8	0.15	0.06	0.78	0.80
PENILAI 9	-0.08	0.07	0.88	0.82
PENILAI 3	-0.10	0.07	0.89	0.71
PENILAI 30	-0.15	0.08	1.16	0.97
PENILAI 20	-0.16	0.08	2.34**	1.81**
PENILAI 4	-0.24	0.07	1.30	0.91
PENILAI 24	-0.28	0.07	1.48	1.12
PENILAI 11	-0.32	0.07	1.17	0.92
PENILAI 1	-0.33	0.08	1.04	0.89
PENILAI 15	-0.60	0.07	0.88	0.83
PENILAI 19	-0.64	0.09	1.56**	1.44

*Nota.* Nilai khi-kuasadua: 545.7 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Sosioemosi 1.* Bagi ukuran ketegasan, Jadual 5.26 menunjukkan statistik pengukuran faset penilai bagi domain sosioemosi IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.86 logit (SE= 0.06) bagi penilai 13 iaitu penilai yang pemurah (*lenient*) sehingga 0.50 logit (SE= 0.04) bagi penilai 4 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.36. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain sosioemosi IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.04 sehingga 0.07.

Bagi perbezaan ketegasan, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 956.4$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain sosioemosi, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat dua orang penilai iaitu penilai 19 (infit MNSQ =1.87, outfit MNSQ = 1.70) dan penilai 23 (infit MNSQ =1.86, outfit MNSQ = 1.73) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain sosioemosi setiap kanak-kanak secara konsisten.

Jadual 5.26

Laporan ukuran ketegasan penilai bagi domain sosioemosi IPHKT 1 menurut urutan ukuran ketegasan

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 4	0.50	0.04	0.87	0.72
PENILAI 27	0.44	0.06	1.01	0.85
PENILAI 3	0.38	0.04	0.94	0.95
PENILAI 8	0.35	0.04	0.80	0.97
PENILAI 9	0.34	0.04	0.87	1.00
PENILAI 32	0.21	0.04	1.17	1.20
PENILAI 31	0.19	0.04	1.16	1.23
PENILAI 25	0.16	0.05	1.12	0.97
PENILAI 26	0.10	0.07	0.92	0.83
PENILAI 23	0.10	0.05	1.86**	1.73**
PENILAI 29	-0.24	0.04	1.04	0.79
PENILAI 30	-0.25	0.04	1.07	0.84
PENILAI 19	-0.27	0.05	1.87**	1.70**
PENILAI 24	-0.35	0.06	0.71	0.58
PENILAI 22	-0.80	0.06	0.84	0.88
PENILAI 13	-0.86	0.06	0.84	0.87

Nota. Nilai khi-kuasadua: 956.4 ; df: 15; Signifikan (kebarangkalian) : 0.00

Domain Sosioemosi 2. Bagi ukuran ketegasan Jadual 5.27 menunjukkan statistik pengukuran faset penilai bagi domain sosioemosi IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -1.21 logit (SE= 0.13) bagi penilai 20 iaitu penilai yang pemurah (*lenient*) sehingga 1.46 logit (SE= 0.10) bagi penilai 26 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 2.67 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain sosioemosi IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.06 sehingga 0.13.

Bagi perbezaan ketegasan, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 1250.9$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain sosioemosi, nilai infit dan outfit

MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat dua orang penilai iaitu penilai 32 (infit MNSQ =1.56, outfit MNSQ = 1.88) dan penilai 20 (infit MNSQ =1.96) yang mempunyai nilai infit dan outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain sosioemosi setiap kanak-kanak secara konsisten.

Jadual 5.27

*Laporan ukuran ketegasan penilai bagi domain sosioemosi IPHKT 2 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 26	1.46	0.10	0.69	0.72
PENILAI 31	1.34	0.06	1.10	0.95
PENILAI 27	1.25	0.11	0.82	0.96
PENILAI 32	0.42	0.08	1.56**	1.88**
PENILAI 9	0.28	0.06	0.73	0.96
PENILAI 29	0.17	0.10	1.19	1.45
PENILAI 1	-0.04	0.07	0.90	0.81
PENILAI 24	-0.13	0.08	0.91	0.78
PENILAI 3	-0.24	0.06	0.93	0.88
PENILAI 30	-0.33	0.12	1.26	1.85
PENILAI 8	-0.35	0.07	0.87	0.89
PENILAI 4	-0.37	0.06	1.20	1.09
PENILAI 11	-0.53	0.07	1.00	0.92
PENILAI 15	-0.67	0.08	0.88	0.84
PENILAI 19	-1.05	0.12	1.05	1.06
PENILAI 20	-1.21	0.13	1.96**	1.31

*Nota.* Nilai khi-kuasadua: 1250.9 ; df: 15; Signifikan (kebarangkalian) : 0.00

*Domain Rohani 1.* Bagi ukuran ketegasan, Jadual 5.28 menunjukkan statistik pengukuran faset penilai bagi domain rohani IPHKT 1 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.90 logit (SE= 0.07) bagi penilai 13 iaitu penilai yang pemurah (*lenient*) sehingga 0.35 logit (SE= 0.06) bagi penilai 25 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.25 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat

piawai bagi penilaian penilai terhadap domain rohani IPHKT 1 kanak-kanak adalah berada pada julat di antara 0.05 sehingga 0.09.

Bagi perbezaan ketegasan, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 405.9$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain sosioemosi, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat dua orang penilai iaitu penilai 19 (outfit MNSQ = 1.60) dan penilai 23 (outfit MNSQ = 1.57) yang mempunyai nilai outfit MNSQ yang melebihi 1.5. Empat belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain rohani setiap kanak-kanak secara konsisten.

Jadual 5.28

*Laporan ukuran ketegasan penilai bagi domain rohani IPHKT 1 menurut urutan ukuran ketegasan*

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 25	0.35	0.06	1.23	1.33
PENILAI 3	0.28	0.05	0.86	1.02
PENILAI 27	0.27	0.09	1.13	1.13
PENILAI 9	0.26	0.06	1.09	1.10
PENILAI 4	0.21	0.05	1.05	1.14
PENILAI 31	0.19	0.05	1.07	1.07
PENILAI 26	0.13	0.09	1.22	1.46
PENILAI 8	0.12	0.06	1.05	1.06
PENILAI 32	0.05	0.05	1.03	0.97
PENILAI 24	0.00	0.07	0.85	0.79
PENILAI 23	-0.03	0.07	1.49	1.57**
PENILAI 29	-0.05	0.05	0.85	0.73
PENILAI 30	-0.07	0.05	0.82	0.64
PENILAI 19	-0.08	0.07	1.37	1.60**
PENILAI 22	-0.71	0.07	0.98	0.96
PENILAI 13	-0.90	0.07	1.08	1.03

*Nota.* Nilai khi-kuasadua: 405.9 ;  $df$ : 15; Signifikan (kebarangkalian) : 0.00

*Domain Rohani 2.* Bagi ukuran Ketegasan, Jadual 5.29 menunjukkan statistik pengukuran faset penilai bagi domain rohani IPHKT 2 terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai infit dan outfit MNSQ. Ketegasan penilai berada pada julat -0.67 logit (SE= 0.09) bagi penilai 15 iaitu penilai yang pemurah (*lenient*) sehingga 0.69 logit (SE= 0.11) bagi penilai 26 iaitu penilai yang tegas. Perbezaan ketegasan di antara kedua penilai ini adalah 1.36 logit. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan. Nilai ralat piawai bagi penilaian penilai terhadap domain rohani IPHKT 2 kanak-kanak adalah berada pada julat di antara 0.07 sehingga 0.13.

Bagi Perbezaan ketegasan, hasil ujian khi kuasa dua menunjukkan terdapat perbezaan yang signifikan ( $\chi^2 = 307.0$ ,  $df = 15$ ,  $p < 0.01$ ) antara penilai-penilai. Ini menunjukkan bahawa penilai-penilai ini menilai dengan ketegasan yang berbeza.

Konsistensi penilai pula, nilai infit dan outfit MNSQ bagi setiap penilai dalam menilai domain pembelajaran kanak-kanak adalah menggambarkan konsistensi penilai dalam membuat penilaian. Dalam menilai domain sosioemosi, nilai infit dan outfit MNSQ bagi setiap penilai adalah berada dalam julat yang boleh diterima iaitu 0.5 dan 1.5. Terdapat seorang penilai iaitu penilai 20 (infit MNSQ = 1.86) yang mempunyai nilai infit MNSQ yang melebihi 1.5. Lima belas penilai yang lain semuanya berada dalam julat yang boleh diterima dan ini menunjukkan mereka telah menilai domain rohani setiap kanak-kanak secara konsisten.



Jadual 5.29

Laporan ukuran ketegasan penilai bagi domain rohani IPHKT 2 menurut urutan ukuran ketegasan

Penilai	Ukuran Ketegasan	Ralat Piawai (SE)	Infit MNSQ	Outfit MNSQ
PENILAI 26	0.69	0.11	0.86	0.99
PENILAI 27	0.60	0.11	0.87	0.93
PENILAI 31	0.43	0.07	1.04	0.98
PENILAI 1	0.42	0.08	1.00	0.88
PENILAI 4	0.29	0.07	1.14	0.90
PENILAI 24	0.17	0.08	0.87	0.71
PENILAI 8	0.12	0.08	0.84	1.06
PENILAI 9	0.03	0.08	0.93	1.05
PENILAI 32	-0.02	0.08	1.11	1.03
PENILAI 3	-0.08	0.07	1.22	1.16
PENILAI 29	-0.13	0.11	1.04	1.08
PENILAI 20	-0.27	0.11	1.86**	1.41
PENILAI 19	-0.43	0.11	1.45	1.10
PENILAI 30	-0.54	0.13	1.10	1.23
PENILAI 11	-0.62	0.09	1.04	0.96
PENILAI 15	-0.67	0.09	0.87	0.73

Nota. Nilai khi-kuasdua: 307.0 ; df: 15; Signifikan (kebarangkalian) : 0.00

Jadual 5.30 dan 5.31 menunjukkan hasil analisis secara keseluruhan bagi menjawab ketiga-tiga persoalan kajian bagi faset penilai. Hasil analisis ukuran ketegasan menunjukkan bahawa ketegasan setiap penilai adalah berbeza. Ini dibuktikan dengan ujian khikuasdua. Konsistensi penilai dalam membuat penilaian terhadap kanak-kanak dengan menggunakan IPHKT 1 dan 2 menunjukkan bahawa hampir semua penilai menilai secara konsisten bagi domain kognitif, domain bahasa, domain fizikal, domain kreativiti, domain sosioemosi dan domain rohani. Walau bagaimanapun, terdapat penilai yang menilai secara tidak konsisten. Penilai yang terlibat adalah penilai 19, penilai 20, penilai 22, penilai 23 dan penilai 32.

Jadual 5.30

Ringkasan statistik bagi faset penilai dalam mengukur domain dalam IPHKT 1

Domain	Nilai julat ukuran ketegasan	Ujian khi kuasadua	statistik kesepadanan			
			nilai infit		Nilai outfit	
			<0.5	>1.5	<0.5	>1.5
Kognitif	-0.52 – 0.55	signifikan		1 (p19)		
Bahasa	-0.59 – 0.62	signifikan		1 (p19)		1 (p19)
Fizikal	-0.93 – 1.06	signifikan		2 (p22&23)		1 (p22)
Kreativiti	-0.71 – 0.59	signifikan		2 (p19&30)		1 (p22)
Sosioemosi	-0.86 – 0.50	signifikan		2 (p19&p23)		1 (p32)
Rohani	-0.90 – 0.35	signifikan				2 (p19&23)

Jadual 5.31

Ringkasan statistik bagi faset penilai dalam mengukur domain dalam IPHKT 2

Domain	Nilai julat ukuran ketegasan	Ujian khi kuasadua	statistik kesepadanan			
			nilai infit		Nilai outfit	
			<0.5	>1.5	<0.5	>1.5
Kognitif	-1.37 – 1.89	signifikan		3 (p19, p20&p32)		1 (p32)
Bahasa	-0.43 – 0.72	signifikan	1 (p27)	2 (p19&p20)		1 (p20)
Fizikal	-0.96 – 1.17	signifikan		1 (p20)		
Kreativiti	-0.64 – 0.77	signifikan		2 (p19&p20)		1 (p20)
Sosioemosi	-1.21 – 1.46	signifikan		2 (p20&p32)		1 (p32)
Rohani	-0.67 – 0.69	signifikan		1 (p20)		

**Laporan Pengukuran Item.** Bahagian ini akan menerangkan dengan terperinci akan kebolehpercayaan dan kesahan bagi item-item bagi setiap domain yang terdapat dalam IPHKT 1.

*Apakah nilai kebolehpercayaan item dan indeks pengasingan item?* Jadual 5.32 dan Jadual 5.33 menunjukkan nilai kebolehpercayaan dan indeks pengasingan item bagi enam domain yang terdapat dalam IPHKT1 dan 2. Nilai kebolehpercayaan item IPHKT 1 berada di antara 0.96 (domain kreativiti) hingga 0.99 (domain bahasa, sosioemosi dan rohani) dan indeks pengasingan di antara 4.82 hingga 11.93. Nilai kebolehpercayaan item IPHKT 2 berada di antara 0.95 hingga 0.98 (domain kognitif

dan kreativiti) dan indeks pengasingan di antara 4.18 hingga 7.33. Nilai ini boleh diterima berdasarkan model pengukuran Rasch.

Jadual 5.32

*Nilai kebolehpercayaan dan indeks pengasingan bagi item IPHKT 1 berdasarkan domain*

Domain	Kebolehpercayaan	Indeks Pengasingan
Kognitif	0.98	8.00
Bahasa	0.99	11.49
Fizikal	0.97	5.85
Kreativiti	0.96	4.82
Sosioemosi	0.99	10.18
Rohani	0.99	11.93

Jadual 5.33

*Nilai kebolehpercayaan dan indeks pengasingan bagi item IPHKT 2 berdasarkan domain*

Domain	Kebolehpercayaan	Indeks Pengasingan
Kognitif	0.98	7.30
Bahasa	0.95	4.27
Fizikal	0.97	5.77
Kreativiti	0.98	7.33
Sosioemosi	0.95	4.18
Rohani	0.97	5.61

Empat persoalan kajian seterusnya dijawab melalui jadual yang mengandungi *nilai measure*, *standard error*, nilai infit MNSQ, outfit MNSQ dan *point measure correlation* bagi setiap domain dalam IPHKT. Keempat-empat soalan ini akan dijawab serentak mengikut setiap domain yang terdapat dalam IPHKT 1 dan 2 kerana ianya merupakan analisis yang menunjukkan kesahan konstruk. Penentuan item-item sama ada dibuang, diubahsuai atau dikekalkan adalah berdasarkan kepada analisis dan hasil yang diperolehi daripada kelima-lima persoalan kajian ini.

***Di manakah kedudukan kesukaran item-item bagi setiap domain yang diukur? Apakah ketepatan penentukuran yang dibuat? Apakah respons item-item bagi setiap domain sepadan (fit) mengikut jangkaan model pengukuran Rasch? Adakah item-item bagi setiap domain bergerak pada satu arah?***

*Domain Kognitif 1.* Rajah 5.1 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item kognitif bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Lajur pertama adalah skala logit. Lajur kedua adalah lajur yang mempamerkan ketegasan penilai, di mana penilai yang mempunyai nilai logit yang tertinggi adalah penilai yang paling tegas. Setiap tanda asterisk (\*) mewakili seorang penilai. Lajur ketiga adalah lajur keupayaan kanak-kanak. Kanak-kanak yang mempunyai nilai logit yang tertinggi adalah kanak-kanak yang mempunyai keupayaan yang tertinggi. Lajur keempat menunjukkan taburan item mengikut tahap kesukaran. Item yang mempunyai nilai logit tertinggi adalah item yang paling sukar. Bagi item-item domain kognitif IPHKT 1 ini, item 16 adalah yang paling sukar. Lajur yang terakhir menunjukkan skala 5 mata yang digunakan dalam penskoran.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
4	+	+	+	(5)
	.			
	.			
	.			
	*****.			
3	+	+	+	
	*			
	.			
	*.			
	.		16	
	*****.			
	**			
	*		4	
	**			
2	+	+	+	
	*			
	*			
	.			
	**			
	*			
	*			
	.			
	**.			
1	+	+	+	4
	*			
	.			
	*	27		
	*	26		
	.			
	.	8	3	
	.	9	17	
	.	4	9	
*	0	* 29 3 30 31	* 10 20	* 3 *
	.	13 24 25 32	2	
	.	23	21 7	
	.		15 19 6 8	
	.		18	
	.	19 22	11 12 13 14	
	.		1 5	
-1	+	+	+	(1)
Logit	* = 2	-Penilai	-Item	Skala

Rajah 5.1. Pemetaan guru penilai, kanak-kanak dan item bagi domain kognitif IPHKT 1

Jadual 5.34 menunjukkan nilai penentukaran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item kognitif IPHKT 1.

Jadual 5.34

*Laporan pengukuran item-item bagi domain kognitif IPHKT 1*

Item	Penentukaran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
16	2.53	0.09	0.59**	0.78	0.53	0.65
4	2.24	0.08	1.16	1.47	0.35	0.67
3	0.29	0.07	1.66**	1.75**	0.39	0.64
17	0.21	0.07	1.22	1.00	0.67	0.63
9	0.07	0.11	1.31	1.64	0.43	0.60
10	0.00	0.07	0.84	0.74	0.71	0.61
20	-0.01	0.07	1.04	0.92	0.63	0.60
2	-0.13	0.12	0.92	0.74	0.62	0.57
7	-0.23	0.08	0.74	0.61	0.70	0.58
21	-0.24	0.13	0.74	0.73	0.62	0.56
6	-0.26	0.08	0.81	0.67	0.69	0.58
19	-0.26	0.08	1.08	1.31	0.46	0.58
8	-0.29	0.08	0.92	0.76	0.63	0.57
15	-0.31	0.08	0.83	0.67	0.68	0.57
18	-0.40	0.08	1.93**	1.71**	0.45	0.56
14	-0.46	0.09	0.96	0.62	0.66	0.55
12	-0.48	0.14	0.69	0.57	0.66	0.52
13	-0.48	0.14	0.80	0.76	0.63	0.52
11	-0.54	0.14	0.92	0.75	0.56	0.51
5	-0.62	0.15	1.23	0.87	0.48	0.50
1	-0.64	0.09	1.15	1.18	0.42	0.53

*Nota.* Nilai khi-kuasdua : 1818.7; df: 20 ; Signifikan (kebarangkalian) : 0.00

Item 16 pada 2.53 logit (SE = 0.09) adalah item yang paling sukar, manakala item 1 pada -0.64 logit (SE = 0.09) adalah item yang paling mudah. Nilai khi-kuasdua adalah 1818.7, df adalah 20,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain kognitif 1 berada di antara 0.07 hingga 0.15.

Jadual 5.34 menunjukkan julat nilai infit MNSQ berada di antara 0.69 – 1.93 manakala nilai outfit MNSQ berada di antara 0.57 – 1.71. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Hampir kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Namun begitu, item 18 (infit MNSQ = 1.93, outfit MNSQ = 1.71) dan item 3 (infit MNSQ = 1.66, outfit MNSQ = 1.75) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Manakala item 16 (infit MNSQ = 0.59) mempunyai nilai infit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.42 (item 1) dan 0.72 (item 10). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Kognitif 2*. Rajah 5.2 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item kognitif bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item domain kognitif IPHKT 2 ini, item 25 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
8	+	+	+	(5)
	*			
7	+	+	+	
	***			
	*			
6	+	+	+	
	****			
	****			
	**			
5	+	+	+	
	*			
	*			
	**			
	*			
4	+	+	+	
	***			
	**			
	*****			
	***			
	*****		*	
3	+	+	+	
	**			
	*****			---
	*****			
	***			
	*			
2	+	+	+	
	****	26 31	*	
	***		*	
	*	27	**	4
	*		**	
1	+	+	+	
	*	29	*	
	*		**	
	*	32	*	---
	*		**	
* 0 *	*	* 1 30	*	* *
	*	15 24 8	****	3
	*	11 3 9	****	
			*****	
-1	+	+	+	---
		19 4	*	
			*	
		20	*	2
-2	+	+	+	(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.2. Pemetaan guru penilai, kanak-kanak dan item bagi domain kognitif IPHKT 2



Jadual 5.35 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item kognitif IPHKT 2.

Jadual 5.35  
Laporan pengukuran item-item bagi domain kognitif IPHKT 2

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
25	3.29	0.11	1.05	1.38	0.36	0.72
23	1.92	0.12	1.66	1.55	0.50	0.66
22	1.59	0.13	1.15	1.09	0.58	0.65
24	1.47	0.13	1.39	1.24	0.61	0.64
1	1.32	0.13	1.21	1.14	0.46	0.63
2	1.12	0.14	1.13	1.13	0.49	0.62
16	1.12	0.14	1.17	1.13	0.48	0.62
21	0.93	0.14	1.27	1.14	0.55	0.60
17	0.67	0.15	1.54	1.22	0.60	0.59
15	0.52	0.15	1.34	1.28	0.48	0.58
9	0.35	0.16	1.24	1.02	0.50	0.56
12	0.22	0.16	1.12	1.22	0.41	0.55
29	0.11	0.17	1.39	1.25	0.58	0.54
26	-0.30	0.11	0.86	0.81	0.60	0.55
10	-0.35	0.11	0.90	0.76	0.58	0.54
20	-0.47	0.12	1.07	0.94	0.57	0.54
27	-0.48	0.12	0.96	0.86	0.53	0.53
8	-0.56	0.12	0.90	0.73	0.61	0.53
11	-0.63	0.20	0.86	0.53	0.62	0.48
28	-0.67	0.20	0.98	1.13	0.44	0.48
19	-0.69	0.12	1.30	1.75	0.42	0.52
30	-0.72	0.12	0.94	0.80	0.53	0.51
32	-0.72	0.12	1.07	0.83	0.55	0.51
4	-0.79	0.12	0.88	0.68	0.61	0.51
3	-0.80	0.13	1.29	0.97	0.54	0.51
6	-0.85	0.13	0.75	0.61	0.62	0.50
18	-0.85	0.13	0.90	0.87	0.57	0.50
7	-0.87	0.13	0.77	0.67	0.60	0.50
31	-0.88	0.13	1.04	0.80	0.53	0.50
5	-1.03	0.13	0.88	0.71	0.56	0.49
14	-1.27	0.24	0.98	0.44	0.56	0.42
13	-1.68	0.28	0.83	0.30	0.55	0.38

Nota. Nilai khi-kuasadua : 2343.9 ; df: 31 ; Signifikan (kebarangkalian) : 0.00

Item 25 pada 3.29 logit (SE = 0.11) adalah item yang paling sukar, manakala item 13 pada -1.68 logit (SE = 0.28) adalah item yang paling mudah. Nilai khi-kuasadua adalah 2343.9, df adalah 31,  $p < 0.01$  bagi item menunjukkan terdapatnya

perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain kognitif 2 berada di antara 0.11 hingga 0.28.

Jadual 5.35 menunjukkan julat nilai infit MNSQ berada di antara 0.75 – 1.66 manakala nilai outfit MNSQ berada di antara 0.30 – 1.75. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Namun begitu, item 23 (infit MNSQ = 1.66) mempunyai nilai infitMNSQ yang melebihi 1.4. Manakala item 19 (outfit MNSQ = 1.75) mempunyai nilai outfit MNSQ yang melebihi 1.4, item 11 (outfit MNSQ = 0.53), item 14 (outfit MNSQ = 0.44) dan item 13 (outfit MNSQ = 0.30) mempunyai nilai infitMNSQ yang kurang daripada 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.36 (item 25) dan 0.61 (item 4, 8 dan 24). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Bahasa 1.* Rajah 5.3 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item bahasa bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Item yang mempunyai nilai logit tertinggi adalah item yang paling sukar. Bagi item-item domain bahasa IPHKT 1 ini, item 29 adalah yang paling sukar.

logit	+Kanak-kanak	-Penilai	-Item	Skala
4	+	+	+	(5)
3	**	+	29	+
	**			
	***			
	*			
	**			
	*			
2	*	+	+	+
	***			
	*			
	*			
	***			
	*****			
	****			---
	*****			
1	*	+	+	+
	*****		26 30	
	****			
	*****	26 27		4
	***		24 25	
	****	24 31	31	---
	***	32		
	*****	25 8	28 35	
* 0	*****	* 29 30 9	*	* 3 *
		19 3	39	---
	*	23	36	
		13 4	22 27 37 38	2
		22	23	
			32	
-1	+	+	34	+
				---
			33	
-2	+	+	+	(1)
logit	* = 1	-Penilai	-Item	Skala

Rajah 5.3. Pemetaan guru penilai, kanak-kanak dan item bagi domain bahasa IPHKT 1

Jadual 5.36 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item bahasa IPHKT 1.

Jadual 5.36  
Laporan pengukuran item-item bagi domain bahasa IPHKT 1

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
29	3.01	0.12	0.69	0.98	0.48	0.51
26	0.93	0.07	0.92	0.84	0.78	0.68
30	0.86	0.07	0.59	0.66	0.76	0.67
25	0.53	0.07	0.66	0.55	0.75	0.62
24	0.50	0.07	0.70	0.60	0.74	0.62
31	0.43	0.07	0.48	0.45	0.75	0.61
28	0.12	0.05	1.33	1.23	0.45	0.53
35	0.07	0.05	1.43	1.42	0.44	0.52
39	-0.30	0.10	0.68	1.56	0.23	0.46
36	-0.41	0.06	1.33	1.30	0.38	0.43
38	-0.45	0.06	0.99	1.26	0.39	0.42
22	-0.45	0.06	1.08	1.16	0.43	0.42
37	-0.50	0.07	0.92	0.99	0.42	0.41
27	-0.50	0.11	1.19	1.45	0.25	0.42
23	-0.60	0.11	0.90	1.62	0.15	0.40
32	-0.85	0.08	1.11	1.53	0.36	0.34
34	-1.03	0.09	0.88	0.78	0.44	0.31
33	-1.36	0.11	0.68	0.45	0.47	0.26

Nota. Nilai khi-kuasadua : 1697.0 ; df: 17 ; Signifikan (kebarangkalian) : 0.00

Item 29 pada 3.01 logit (SE = 0.12) adalah item yang paling sukar, manakala item 33 pada -1.36 logit (SE = 0.11) adalah item yang paling mudah. Nilai khi-kuasadua adalah 1697.0, df adalah 17,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain bahasa 1 berada di antara 0.05 hingga 0.12.

Jadual 5.36 menunjukkan julat nilai infit MNSQ berada di antara 0.48 – 1.43 manakala nilai outfit MNSQ berada di antara 0.45 – 1.62. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Namun begitu, item 35 (infit MNSQ = 1.43, outfit MNSQ = 1.42) mempunyai nilai infitMNSQ yang melebihi 1.4. Item 30 (infitMNSQ = 0.59) dan item 31 (infitMNSQ = 0.48, outfit MNSQ = 0.45 ) mempunyai nilai infit dan outfit MNSQ yang menghampiri 0.6. Manakala item 23 (outfit MNSQ = 1.62), item 27 (outfit MNSQ = 1.45) dan item 32 (outfit MNSQ = 1.53) mempunyai nilai outfit MNSQ yang melebihi 1.4, item 33 (outfit MNSQ = 0.45) dan item 25 (outfit MNSQ = 0.55) mempunyai nilai outfit MNSQ yang di bawah 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.15 (item 23) dan 0.78 (item 26). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Bahasa 2*. Rajah 5.4 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item bahasa bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item domain bahasa IPHKT 2 ini, item 52 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
5	*** *	+	+	(5)
	*			
	*			
	*			
4	* *	+	+	
	*			
	**			
	**** ****			
3	+	+	+	
	*			
	**** *			
	**			
	**			
2	***** ***** ****	+	+	
	**			
	***** ****			---
	**			
1	*** ** *****	+	+	
	**	26 31	*	4
	**	27	**	
	**	29	***	
	*	29	**	---
* 0	* * * * * * * * * * * *	* 1 20 3 11 24 32 4 8 9 15 30 19	* ***** * ** ** * * * *	* * 3 * * 2 * * * *
				---
-1	+	+	+	
				---
-2	+	+	+	(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.4. Pemetaan guru penilai, kanak-kanak dan item bagi domain bahasa IPHKT 2

Jadual 5.37 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item bahasa IPHKT 2.

Jadual 5.37  
Laporan pengukuran item-item bagi domain bahasa IPHKT 2

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
52	0.98	0.06	1.38	1.41	0.50	0.60
48	0.76	0.06	1.14	1.11	0.55	0.58
41	0.58	0.10	0.90	1.30	0.48	0.55
45	0.47	0.11	1.26	1.26	0.50	0.54
47	0.46	0.06	1.14	1.13	0.53	0.54
55	0.31	0.11	1.02	1.23	0.49	0.52
39	0.28	0.12	1.12	1.26	0.47	0.51
50	0.22	0.07	0.87	0.75	0.60	0.51
37	0.11	0.12	1.14	0.93	0.50	0.49
44	0.10	0.07	0.79	0.78	0.55	0.50
43	0.07	0.07	1.50	1.39	0.46	0.50
42	0.05	0.13	0.72	0.91	0.51	0.48
49	0.02	0.07	1.01	0.94	0.49	0.49
51	0.02	0.07	0.80	0.90	0.49	0.49
33	0.01	0.13	1.03	0.99	0.42	0.47
53	0.01	0.07	1.01	1.05	0.49	0.49
54	0.00	0.07	0.85	0.90	0.51	0.49
46	-0.01	0.07	0.98	0.89	0.54	0.49
35	-0.06	0.08	0.85	0.86	0.52	0.48
36	-0.16	0.14	0.73	0.57	0.54	0.45
38	-0.21	0.14	1.03	1.51	0.43	0.44
40	-0.25	0.14	0.90	0.75	0.47	0.44
34	-0.33	0.15	0.71	1.05	0.48	0.43
58	-0.37	0.08	0.97	1.16	0.41	0.44
59	-0.52	0.09	0.89	0.85	0.46	0.42
56	-0.92	0.19	1.26	0.99	0.37	0.35
57	-1.61	0.25	0.93	0.54	0.33	0.27

Nota. Nilai khi-kuasadua :616.0 ; df: 26 ; Signifikan (kebarangkalian) : 0.00

Item 52 pada 0.98 logit (SE = 0.06) adalah item yang paling sukar, manakala item 57 pada -1.61 logit (SE = 0.25) adalah item yang paling mudah. Nilai khi-kuasadua adalah 616.0, df adalah 26,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain bahasa 2 berada di antara 0.06 hingga 0.25.

Jadual 5.37 menunjukkan julat nilai infit MNSQ berada di antara 0.71 – 1.50 manakala nilai outfit MNSQ berada di antara 0.54 – 1.51. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali item 43 (infit MNSQ = 1.51) mempunyai nilai infitMNSQ yang melebihi 1.4 Manakala item 52 (outfit MNSQ = 1.41) dan item 38 (outfit MNSQ = 1.51) mempunyai nilai outfit MNSQ yang melebihi 1.4, item 36 (outfit MNSQ = 0.57) dan item 57 (outfit MNSQ = 0.54) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.33 (item 57) dan 0.60 (item 60). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Fizikal 1.* Rajah 5.5 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item fizikal bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item domain fizikal IPHKT 1 ini, item 68 adalah yang paling sukar.



Logit	+Kanak-kanak	-Penilai	-Item	Skala
5	*****	+	+	(5)
	*****			
	*			
4	*	+	+	
	*			
	*			
	*			
	*****			
3	**	+	+	
	*			
	*****			
	***			
	*****			
	****			
2	****	+	+	
	*			
	****			
	*****			---
	***			
	***		*	
	*****		*	
1	**	+ 26 27	+ ***	+
	***			4
	*			
	***	31	***	
	*	32	***	
			***	---
		22 29 30	**	
* 0	*	* 13	* *	* 3 *
		3 8 9	***	
			*****	
	*	25	*	---
		19 23 4		
			*****	
		24	*	2
-1	+	+	**	+
			*	---
-2	+	+	+	(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.5. Pemetaan guru penilai, kanak-kanak dan item bagi domain fizikal IPHKT 1

Jadual 5.38 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain fizikal IPHKT 1.

Jadual 5.38  
Laporan pengukuran item-item bagi domain fizikal IPHKT 1

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
68	1.25	0.09	0.59	0.55	0.72	0.62
42	1.12	0.10	1.22	1.22	0.51	0.60
41	1.01	0.10	1.23	1.20	0.50	0.59
50	0.98	0.10	1.51	2.12	0.38	0.58
66	0.95	0.10	1.50	1.23	0.54	0.58
71	0.63	0.11	1.20	1.02	0.49	0.54
64	0.56	0.11	1.14	1.48	0.51	0.53
63	0.55	0.11	1.75	1.79	0.41	0.53
72	0.48	0.11	1.10	0.99	0.49	0.52
61	0.47	0.11	0.65	0.71	0.59	0.52
57	0.40	0.12	0.82	0.74	0.54	0.51
58	0.27	0.12	0.67	0.54	0.60	0.49
70	0.24	0.12	0.78	0.70	0.56	0.48
67	0.22	0.12	1.71	1.15	0.43	0.48
60	0.21	0.13	0.50	0.40	0.61	0.48
62	0.10	0.13	0.77	0.71	0.52	0.46
49	-0.01	0.09	1.34	2.44	0.36	0.56
52	-0.12	0.09	1.07	0.80	0.59	0.54
45	-0.14	0.09	1.24	1.33	0.43	0.54
59	-0.17	0.14	0.47	0.36	0.59	0.42
47	-0.22	0.10	1.36	0.83	0.56	0.53
44	-0.31	0.10	1.09	1.00	0.52	0.52
53	-0.31	0.10	1.18	0.73	0.57	0.52
48	-0.33	0.10	1.10	0.71	0.58	0.52
40	-0.43	0.16	0.93	1.05	0.40	0.39
56	-0.69	0.11	1.09	1.69	0.42	0.47
55	-0.75	0.12	0.82	0.56	0.54	0.47
51	-0.76	0.12	1.11	0.56	0.52	0.46
69	-0.77	0.12	0.93	0.58	0.53	0.46
65	-0.79	0.12	1.03	1.29	0.49	0.46
46	-1.03	0.13	0.71	0.62	0.52	0.43
54	-1.07	0.13	1.24	0.89	0.45	0.42
43	-1.53	0.16	0.79	1.39	0.35	0.36

Nota. Nilai khi-kuasadua : 1205.8 ; df: 32 ; Signifikan (kebarangkalian) : 0.00

Item 68 pada 1.25 logit (SE = 0.09) adalah item yang paling sukar, manakala item 43 pada -1.53 logit (SE = 0.16) adalah item yang paling mudah. Nilai khi-kuasadua adalah 1205.8, df adalah 32,  $p < 0.01$  bagi item menunjukkan terdapatnya

perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain fizikal 1 berada di antara 0.09 hingga 0.16.

Jadual 5.38 menunjukkan julat nilai infit MNSQ berada di antara 0.47 – 1.75 manakala nilai outfit MNSQ berada di antara 0.40 – 2.44. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali, item 50 (infit MNSQ = 1.51, outfit MNSQ = 2.12) dan item 63 (infit MNSQ = 1.75, outfit MNSQ = 1.79) mempunyai nilai infit dan outfitMNSQ yang melebihi 1.4. Item 66 (infit MNSQ = 1.50), item 64 (outfit MNSQ = 1.48), item 49 (outfit MNSQ = 2.44) dan item 56 (outfit MNSQ = 1.69) mempunyai nilai infit atau outfitMNSQ yang melebihi 1.4. Manakala item 68 (infit MNSQ = 0.59, outfit MNSQ = 0.55), item 60 (infit MNSQ = 0.50, outfit MNSQ = 0.40) dan item 59 (infit MNSQ = 0.47, outfit MNSQ = 0.36) mempunyai nilai infit dan outfit MNSQ yang menghampiri 0.6. Item 58 (outfit MNSQ = 0.54), item 55 (outfit MNSQ = 0.56), item 51 (outfit MNSQ = 0.56) dan item 69 (outfit MNSQ = 0.58) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.35 (item 43) dan 0.72 (item 68). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Fizikal 2.* Rajah 5.6 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item fizikal bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item domain fizikal IPHKT 2

ini, item 60 adalah yang paling sukar. Lajur yang terakhir menunjukkan skala 5 mata yang digunakan dalam penskoran.

Ukuran	Kanak-kanak	Penilai	Item	Skala
6	***			(5)
	***			
5	**			
	*			
	****			
4	**			
	*****			
	*****			
	*			
	***			
3	***			
	***			
	*****			
	*****			
	*****			
	*****			
2	**			---
	*			
	****			
	***		**	
	*****	26 27	**	
1		31	* 4	
			**	
			*	
	*	29 32	**	---
*	0	* 30	* *	* *
		1 15 3 8	**	3
		11 24 4	***	
		9	*	
			*****	---
-1		+ 19 20	+ *	+ 2
			*	
			*	
-2				---
				(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.6. Pemetaan guru penilai, kanak-kanak dan item bagi domain fizikal IPHKT 2

Jadual 5.39 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain fizikal IPHKT 2.

Jadual 5.39  
Laporan pengukuran item-item bagi domain fizikal IPHKT 2

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
60	1.35	0.07	1.11	1.16	0.46	0.58
61	1.34	0.07	1.02	1.04	0.43	0.58
82	1.19	0.11	0.88	1.01	0.41	0.55
81	1.11	0.12	1.19	1.04	0.51	0.51
77	0.97	0.13	1.67	1.29	0.48	0.49
83	0.85	0.13	1.12	0.98	0.49	0.47
72	0.77	0.13	1.11	0.87	0.51	0.47
75	0.73	0.13	1.02	1.15	0.45	0.46
78	0.54	0.14	0.74	0.68	0.52	0.44
63	0.46	0.15	0.81	1.09	0.41	0.43
79	0.05	0.17	1.32	1.00	0.39	0.38
71	-0.16	0.18	0.91	0.67	0.44	0.35
64	-0.25	0.11	1.89	1.43	0.37	0.41
84	-0.33	0.19	1.34	1.35	0.29	0.33
73	-0.36	0.19	0.84	0.58	0.44	0.33
76	-0.36	0.19	0.79	0.57	0.43	0.33
70	-0.58	0.12	1.35	0.99	0.40	0.38
67	-0.65	0.12	1.04	1.17	0.41	0.37
69	-0.65	0.12	1.06	1.03	0.41	0.37
65	-0.67	0.12	0.95	0.78	0.42	0.37
74	-0.70	0.22	0.99	0.63	0.38	0.29
68	-0.73	0.12	0.89	0.73	0.44	0.36
80	-1.00	0.14	0.97	0.80	0.40	0.33
66	-1.28	0.15	0.91	0.70	0.37	0.30
62	-1.63	0.18	1.22	0.65	0.32	0.26

Nota. Nilai khi-kuasadua : 1339.6 ; df: 24 ; Signifikan (kebarangkalian) : 0.00

Item 60 pada 1.35 logit (SE = 0.07) adalah item yang paling sukar, manakala item 62 pada -1.63 logit (SE = 0.18) adalah item yang paling mudah. Nilai khi-kuasadua adalah 1339.6, df adalah 24,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain fizikal 2 berada di antara 0.07 hingga 0.22.

Jadual 5.39 menunjukkan julat nilai infit MNSQ berada di antara 0.74 – 1.89 manakala nilai outfit MNSQ berada di antara 0.57 – 1.43. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali, item 64 (infit MNSQ = 1.89, outfit MNSQ = 1.43) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Item 77 (infit MNSQ = 1.67) mempunyai nilai infit MNSQ yang melebihi 1.4. Manakala item 73 (outfit MNSQ = 0.58) dan item 76 (outfit MNSQ = 0.57) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.29 (item 84) dan 0.52 (item 78). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Kreativiti 1.* Rajah 5.7 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item kreativiti bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item kreativiti IPHKT 1 ini, item 90 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item		Skala
+ 4 +	*****.	+			+ (5) +
	*				
+ 3 +	.	+			+ +
	*				
	**.				
	.				
	*.				
	*				
+ 2 +	*	+			+ +
	***.				
	.				
	***				
	*.				
	***				
	*.				
	.				
+ 1 +	**	+			+ +
	*		90		
	*.		81	82	4
	*.		79	83	
	*	26 32	76		
	*	27 31	92		---
		8			
	.	30 9	74	75	
* 0 *	.	* 25	* 86		* *
	.	24 29 3	77		3
		23 4	78 87 88		
			85 89		
		13	73		---
		22	80 91		
		19			
+ -1 +		+	84		+ 2 +
					---
+ -2 +		+			+ (1) +
Logit	* = 2	-Penilai	-Item		Skala

Rajah 5.7. Pemetaan guru penilai, kanak-kanak dan item domain kreativiti IPHKT 1

Jadual 5.40 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain kreativiti IPHKT 1.

Jadual 5.40  
Laporan pengukuran item-item bagi domain kreativiti IPHKT 1

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
90	0.90	0.10	1.03	1.02	0.59	0.59
81	0.72	0.10	1.02	1.07	0.56	0.56
82	0.70	0.10	0.86	0.87	0.58	0.55
83	0.62	0.10	0.99	1.11	0.54	0.54
79	0.58	0.10	1.20	1.31	0.50	0.53
76	0.48	0.11	0.67	0.74	0.59	0.52
92	0.38	0.06	1.03	1.48	0.46	0.50
74	0.13	0.12	0.79	0.77	0.50	0.46
75	0.12	0.12	0.79	0.79	0.51	0.46
86	-0.04	0.13	1.67	2.94	0.27	0.43
77	-0.17	0.08	1.04	0.89	0.44	0.41
87	-0.20	0.08	0.92	0.84	0.44	0.41
88	-0.27	0.08	1.05	0.85	0.43	0.40
78	-0.28	0.08	1.17	0.97	0.41	0.40
85	-0.41	0.08	0.94	0.99	0.40	0.38
89	-0.41	0.08	1.06	0.89	0.40	0.38
73	-0.54	0.09	1.33	1.33	0.34	0.36
91	-0.60	0.16	1.23	1.24	0.31	0.35
80	-0.68	0.09	1.33	0.95	0.35	0.34
84	-1.02	0.20	0.65	0.64	0.37	0.29

Nota. Nilai khi-kuasdua : 524.9 ; df: 19 ; Signifikan (kebarangkalian) : 0.00

Item 90 pada 0.90 logit (SE = 0.10) adalah item yang paling sukar, manakala item 84 pada -1.02 logit (SE = 0.20) adalah item yang paling mudah. Nilai khi-kuasdua adalah 524.9, df adalah 19,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain kreativiti 1 berada di antara 0.06 hingga 0.20.



Jadual 5.40 menunjukkan julat nilai infit MNSQ berada di antara 0.65 – 1.67 manakala nilai outfit MNSQ berada di antara 0.64 – 2.94. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali, item 86 (infit MNSQ = 1.67, outfit MNSQ = 2.94) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4 Manakala item 92 (outfit MNSQ = 1.48) mempunyai nilai outfit MNSQ yang melebihi 1.4.

Nilai *point measure correlation* pula berada di antara nilai 0.27 (item 86) dan 0.59 (item 76 dan 90). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Kreativiti 2*. Rajah 5.8 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item kreativiti bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item kreativiti IPHKT 2 ini, item 99 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai		-Item	Skala
+ 5 +		+		+	+ (5) +
	*				
	*				
	**				
+ 4 +	+ ****	+		+	+ +
	**				
	**				
	**				
	****			*	
	**				
+ 3 +	+ **	+		+	+ +
	***				
	*****				
	*****				
	****				
	**				
+ 2 +	+ ****	+		+	+ +
	***				---
	*****				
	*				
	*				
	****				
+ 1 +	+ ****	+		+	+ 4 +
	***			*	
	*	26 27 31		*	
	*			*	
	**			*	---
		29 32		**	
		8		**	
* 0 *	* *	* *		* ***	* 3 *
		20 3 30 9		*	
		1 11 24 4		*****	
				*****	---
		15 19		**	
+ -1 +		+		+	+ 2 +
				*	
					---
+ -2 +		+		+	+ (1) +
Logit	* = 1	-Penilai		* = 1	Skala

Rajah 5.8. Pemetaan guru penilai, kanak-kanak dan item domain kreativiti IPHKT 2

Jadual 5.41 menunjukkan nilai penentukaran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain kreativiti IPHKT 2.

Jadual 5.41  
Laporan pengukuran item-item bagi domain kreativiti IPHKT 2

Item	Penentukaran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
99	3.25	0.09	0.74	0.98	0.45	0.68
95	0.93	0.11	1.31	1.01	0.59	0.65
96	0.64	0.11	0.92	0.71	0.65	0.62
97	0.37	0.12	0.84	0.87	0.59	0.58
89	0.30	0.07	1.37	1.27	0.53	0.60
110	0.23	0.13	1.12	1.36	0.36	0.56
88	0.20	0.07	1.48	1.36	0.49	0.56
105	0.15	0.13	1.27	1.16	0.45	0.59
106	-0.02	0.08	0.83	0.73	0.62	0.55
90	-0.04	0.08	1.40	1.17	0.49	0.56
98	-0.07	0.14	1.20	0.91	0.56	0.56
101	-0.08	0.08	1.07	0.98	0.55	0.51
109	-0.23	0.15	0.96	0.80	0.54	0.55
104	-0.24	0.08	0.76	0.65	0.61	0.50
93	-0.26	0.08	0.93	0.77	0.62	0.53
107	-0.34	0.15	0.92	1.86	0.21	0.53
91	-0.35	0.08	1.76	1.46	0.41	0.48
92	-0.35	0.08	0.91	0.70	0.63	0.51
103	-0.36	0.08	0.76	0.56	0.63	0.51
85	-0.39	0.08	1.21	1.25	0.48	0.51
94	-0.39	0.08	0.99	0.82	0.59	0.51
108	-0.39	0.15	0.94	0.78	0.51	0.51
102	-0.41	0.09	0.79	0.63	0.63	0.47
87	-0.53	0.09	0.82	1.09	0.48	0.51
100	-0.53	0.09	1.01	0.90	0.52	0.49
86	-1.11	0.11	1.02	0.71	0.42	0.41

Nota. Nilai khi-kuasadua : 1720.1 ; df: 25 ; Signifikan (kebarangkalian) : 0.00

Item 99 pada 3.25 logit (SE = 0.09) adalah item yang paling sukar, manakala item 86 pada -1.11 logit (SE = 0.11) adalah item yang paling mudah. Nilai khi-kuasadua adalah 1720.1, df adalah 25,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain kreativiti 2 berada di antara 0.07 hingga 0.16.

Jadual 5.41 menunjukkan julat nilai infit MNSQ berada di antara 0.74 – 1.76 manakala nilai outfit MNSQ berada di antara 0.56 – 1.86. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali, item 91 (infit MNSQ = 1.76, outfit MNSQ = 1.46) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Item 88 (infit MNSQ = 1.48) mempunyai nilai infitMNSQ yang melebihi 1.4. Manakala item 107 (outfit MNSQ = 1.86) mempunyai nilai outfit MNSQ yang melebihi 1.4 dan item 103 (outfit MNSQ = 0.56) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.21 (item 107) dan 0.65 (item 96). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Sosioemosi 1*. Rajah 5.9 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item sosioemosi bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item sosioemosi IPHKT 1 ini, item 127 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
3	+	+	+ *	+ (5)
			*	
	*			
	*		*	
	***			
	***			
	*****			
2	+ **	+	+	+
	*			
	*			
	**		*	
	*****		*	
	*****			
	*****			
	*			---
	****		*	
	*		*	
1	+ ***	+	+	+
	*****			
	***			
	*****			
	*****			4
	**	4		
	**	27 3		
	*	8 9	**	---
	****	25 31 32		
	**	23 26	**	
* 0	* **	*	* *****	* 3 *
	**	29 30	*****	
	*	19		---
	**	24	****	
	*		****	
			*	
			**	2
		22	****	
		13		
-1	+	+	+ ***	+
			*	---
-2	+	+	+	+
				(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.9. Pemetaan guru penilai, kanak-kanak dan item domain sosioemosi IPHKT 1

Jadual 5.42

Laporan pengukuran item-item bagi domain sosioemosi IPHKT 1

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
127	2.97	0.12	0.50	0.81	0.37	0.35
124	2.86	0.11	0.65	0.97	0.31	0.38
95	2.37	0.9	0.92	1.30	0.42	0.48
94	1.74	0.07	0.56	0.85	0.37	0.60
137	1.58	0.07	0.75	0.88	0.41	0.63
131	1.17	0.07	0.82	1.25	0.19	0.67
134	1.06	0.07	0.86	1.15	0.18	0.67
128	0.30	0.08	1.26	1.06	0.63	0.62
125	0.29	0.08	0.84	0.86	0.56	0.61
120	0.11	0.09	0.98	0.80	0.57	0.59
96	0.07	0.05	1.35	1.35	0.55	0.57
129	0.04	0.09	1.34	1.72	0.32	0.57
132	0.01	0.05	1.24	1.07	0.56	0.56
106	0.00	0.09	1.22	0.95	0.63	0.57
102	-0.02	0.05	1.37	1.21	0.55	0.55
104	-0.02	0.05	1.20	1.15	0.53	0.55
119	-0.05	0.09	1.57	1.18	0.57	0.56
100	-0.06	0.09	1.03	0.77	0.64	0.56
103	-0.12	0.05	1.26	1.26	0.50	0.53
108	-0.14	0.09	0.95	0.73	0.67	0.54
117	-0.15	0.09	1.14	0.81	0.62	0.54
123	-0.20	0.10	0.76	0.77	0.51	0.53
118	-0.21	0.10	1.09	0.74	0.63	0.53
93	-0.24	0.10	1.58	1.40	0.45	0.52
99	-0.24	0.10	1.09	1.09	0.48	0.52
109	-0.24	0.10	1.07	0.99	0.59	0.52
111	-0.37	0.10	1.00	0.69	0.64	0.49
101	-0.38	0.10	0.83	0.65	0.65	0.49
113	-0.40	0.06	1.36	1.27	0.43	0.48
110	-0.44	0.06	0.83	0.76	0.56	0.47
126	-0.45	0.06	0.99	0.85	0.53	0.47
105	-0.47	0.11	1.14	0.79	0.58	0.47
107	-0.48	0.11	0.75	0.56	0.64	0.47
112	-0.54	0.11	0.56	0.44	0.67	0.46
97	-0.61	0.12	0.81	0.73	0.52	0.45
130	-0.71	0.07	1.22	1.15	0.38	0.42
136	-0.75	0.07	1.02	0.94	0.43	0.41
133	-0.77	0.07	0.90	0.82	0.47	0.41
121	-0.78	0.13	1.40	1.04	0.51	0.41
98	-0.79	0.13	1.23	0.83	0.48	0.41
115	-0.82	0.13	1.18	0.62	0.55	0.40
135	-0.96	0.08	0.91	0.82	0.42	0.37
114	-0.96	0.14	1.00	0.98	0.41	0.38
116	-1.03	0.15	1.40	0.62	0.54	0.37
122	-1.17	0.09	1.07	0.85	0.39	0.33

Nota. Nilai khi-kuasadua : 4658.3 ; df: 44 ; Signifikan (kebarangkalian) : 0.00

Jadual 5.42 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain sosioemosi IPHKT 1.

Item 127 pada 2.97 logit (SE = 0.12) adalah item yang paling sukar, manakala item 122 pada -1.17 logit (SE = 0.09) adalah item yang paling mudah. Nilai khikwasadua adalah 4658.3, df adalah 44,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain sosioemosi 1 berada di antara 0.05 hingga 0.15.

Jadual 5.42 menunjukkan julat nilai infit MNSQ berada di antara 0.50 – 1.58 manakala nilai outfit MNSQ berada di antara 0.44 – 1.72. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali item 93 (infit MNSQ = 1.58) dan item 119 (infit MNSQ = 1.57) mempunyai nilai infitMNSQ yang melebihi 1.4 Manakala item 129 (outfit MNSQ = 1.72) mempunyai nilai outfit MNSQ yang melebihi 1.4. Item 112 (infit MNSQ = 0.56, outfit MNSQ = 0.44) mempunyai nilai infit dan outfit MNSQ yang bawah 0.6. Manakala item 94 (infit MNSQ = 0.56) dan item 127 (infit MNSQ = 0.50) mempunyai nilai infit MNSQ yang menghampiri 0.6, item 107 (outfit MNSQ = 0.56) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.18 (item 134) dan 0.67 (item 108 dan 112). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Sosioemosi 2.* Rajah 5.10 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item sosioemosi bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item sosioemosi IPHKT 2 ini, item 146 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
6	*****	+	+	(5)
	*			
5	+	+	+	
	*			
	*			
4	***	+	+	
	**			
	*			
	*			
	*			
3	+	+	+	
	*			
	*****			
	**			
	****			
	****			
2	*****	+	+	
	*****			---
	*****			
	*****	26		
	*	27 31		
	*****		*	
1	+	+	+	4
	*****		**	
	*		*	
	*	32	*	
	***	9	*****	
	*	29	*****	---
* 0 *	*	* 1	* ***	* --- *
		24	****	
		3 30 4 8	***	3
		11	**	---
		15	*	
	*			
-1	+	+	+	2
		19	***	
		20	**	
				---
-2	+	+	+	(1)
Logit	* = 1	-Penilai	* = 1	Skala

Rajah 5.10. Pemetaan guru penilai, kanak-kanak dan item domain sosioemosi IPHKT 2



Jadual 5.43

Laporan pengukuran item-item bagi domain sosioemosi IPHKT 2

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
146	1.15	0.10	0.87	0.79	0.61	0.60
126	1.03	0.10	1.49	1.73	0.43	0.58
147	0.98	0.11	0.76	0.83	0.61	0.58
131	0.67	0.11	0.84	1.04	0.55	0.55
130	0.43	0.12	0.72	0.94	0.54	0.52
119	0.40	0.12	1.09	1.33	0.49	0.52
120	0.40	0.12	0.95	0.98	0.54	0.52
123	0.39	0.12	1.04	0.94	0.49	0.52
111	0.38	0.08	1.45	1.50	0.41	0.51
116	0.36	0.12	0.78	0.73	0.56	0.51
127	0.36	0.12	0.66	0.83	0.55	0.51
114	0.32	0.08	1.33	1.38	0.42	0.50
134	0.28	0.13	0.78	0.84	0.53	0.50
133	0.26	0.13	0.89	0.91	0.49	0.50
139	0.26	0.13	0.74	0.94	0.55	0.50
138	0.23	0.13	0.80	0.97	0.55	0.50
129	0.21	0.13	0.59	0.81	0.55	0.50
125	0.15	0.13	0.90	0.94	0.53	0.49
132	0.15	0.13	0.88	0.75	0.55	0.49
135	0.15	0.13	0.72	0.85	0.56	0.49
148	0.11	0.13	0.85	0.78	0.51	0.49
118	0.06	0.13	1.32	1.09	0.45	0.48
113	-0.05	0.14	0.99	0.77	0.52	0.47
117	-0.07	0.14	1.37	0.90	0.51	0.47
122	-0.09	0.14	0.94	0.98	0.44	0.46
112	-0.13	0.14	0.74	0.72	0.55	0.46
140	-0.13	0.14	0.82	0.92	0.51	0.46
143	-0.24	0.15	0.72	1.24	0.49	0.45
137	-0.26	0.15	1.28	1.61	0.44	0.45
149	-0.31	0.15	0.83	1.01	0.48	0.44
121	-0.35	0.15	0.92	0.96	0.39	0.44
124	-0.42	0.16	0.88	0.76	0.47	0.43
136	-0.42	0.16	1.21	0.94	0.46	0.43
128	-0.85	0.11	1.22	1.07	0.36	0.38
145	-0.94	0.11	1.17	1.05	0.36	0.38
115	-1.01	0.19	1.37	1.47	0.31	0.37
141	-1.03	0.12	1.65	2.11	0.31	0.37
142	-1.19	0.12	0.97	0.80	0.39	0.35
144	-1.20	0.20	0.96	0.73	0.37	0.35

Nota. Nilai khi-kuasadua : 837.4 ; df: 38 ; Signifikan (kebarangkalian) : 0.00

Jadual 5.43 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain sosioemosi IPHKT 2.

Item 146 pada 1.15 logit (SE = 0.10) adalah item yang paling sukar, manakala item 144 pada -1.20 logit (SE = 0.20) adalah item yang paling mudah. Nilai khikuasdua adalah 837.4, df adalah 38,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain sosioemosi 2 berada di antara 0.08 hingga 0.20.

Jadual 5.43 menunjukkan julat nilai infit MNSQ berada di antara 0.59 – 1.65 manakala nilai outfit MNSQ berada di antara 0.72 – 2.11. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali item 141 (infit MNSQ = 1.65, outfit MNSQ = 2.11), item 126 (infit MNSQ = 1.49, outfit MNSQ = 1.73) dan item 111 (infit MNSQ = 1.45, outfit MNSQ = 1.50) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Manakala item 137 (outfit MNSQ = 1.61) dan item 115 (outfit MNSQ = 1.47) mempunyai nilai outfit MNSQ yang melebihi 1.4. Item 129 (infit MNSQ = 0.59) mempunyai nilai infit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.31 (item 115 dan 141) dan 0.61 (item 146 dan 147). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Rohani 1.* Rajah 5.11 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item rohani bagi IPHKT 1 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item rohani IPHKT 1 ini, item 140 adalah yang paling sukar.

Logit	+Kanak-kanak	-Penilai	-Item	Skala
5	+	+	+	(5)
			*	
4	+	+	+	+
3	+	+	+	+
	*		*	
	*.		*	
2	+	+	+	+
	**		*	
	**		*	
	***		*	
	****		*	
	****.		*	
	**			---
	**			
1	+	+	+	+
	*.		*	
	*****.		*	
	*		*	4
	**.			
	*			
	**.	25 27 3 4 9		---
	**.	26 31 8	***	
* 0	***	* 23 24 29 32	* **	* 3 *
	.	19 30	*	
	*		*	---
			***	
			****	
		22	***	2
		13	***	
-1	+	+	+	+
			**	
			**	---
			*	
			*	
-2	+	+	+	(1)
Logit	* = 2	-Penilai	* = 1	Skala

Rajah 5.11. Pemetaan guru penilai, kanak-kanak dan item bagi domain rohani IPHKT 1

Jadual 5.44 menunjukkan nilai penentukaran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain rohani IPHKT 1.

Jadual 5.44  
Laporan pengukuran item-item bagi domain rohani IPHKT 1

Item	Penentukaran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
140	4.39	0.32	1.70	2.54	0.0	0.15
138	2.85	0.13	0.54	0.88	0.44	0.36
150	2.28	0.09	0.89	1.00	0.51	0.46
142	2.07	0.09	1.25	1.19	0.55	0.50
146	1.57	0.07	0.38	0.47	0.49	0.56
147	1.40	0.07	0.51	0.51	0.53	0.58
164	0.80	0.07	0.45	0.54	0.49	0.60
139	0.72	0.07	0.62	0.70	0.45	0.59
158	0.20	0.05	1.51	1.36	0.59	0.57
165	0.18	0.08	0.69	0.78	0.58	0.55
157	0.13	0.05	1.53	1.35	0.58	0.56
145	-0.03	0.09	0.99	0.97	0.57	0.52
143	-0.06	0.09	0.64	0.68	0.65	0.51
166	-0.09	0.09	0.67	0.70	0.49	0.51
167	-0.27	0.09	0.88	0.86	0.45	0.48
169	-0.37	0.10	0.72	0.84	0.53	0.46
144	-0.50	0.10	0.50	0.45	0.69	0.44
170	-0.50	0.10	0.92	1.09	0.47	0.44
152	-0.52	0.11	0.69	0.74	0.52	0.42
168	-0.59	0.11	1.01	0.85	0.51	0.42
149	-0.60	0.11	1.34	0.96	0.51	0.42
171	-0.62	0.11	0.77	0.61	0.55	0.41
154	-0.65	0.06	1.21	1.29	0.33	0.42
160	-0.65	0.06	1.46	1.61	0.28	0.42
153	-0.72	0.12	0.66	0.70	0.48	0.40
141	-0.82	0.12	0.78	0.57	0.52	0.38
163	-0.82	0.12	1.01	0.67	0.52	0.38
148	-0.88	0.13	0.91	1.06	0.36	0.37
162	-1.08	0.14	0.62	0.53	0.52	0.33
161	-1.09	0.08	1.43	1.12	0.35	0.34
151	-1.23	0.09	1.69	1.26	0.31	0.31
156	-1.26	0.09	2.06	1.46	0.29	0.31
159	-1.60	0.11	1.26	1.16	0.21	0.25
155	-1.67	0.11	1.01	0.91	0.24	0.24

Nota. Nilai khi-kuasadua : 4615.8 ; df:33 ; Signifikan (kebarangkalian) : 0.00

Item 140 pada 4.39 logit (SE = 0.32) adalah item yang paling sukar, manakala item 155 pada -1.67 logit (SE = 0.11) adalah item yang paling mudah. Nilai khi-kuasadua adalah 4615.8, df adalah 33,  $p < 0.01$  bagi item menunjukkan terdapatnya

perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain rohani 1 berada di antara 0.05 hingga 0.32.

Jadual 5.44 menunjukkan julat nilai infit MNSQ berada di antara 0.38 – 2.06 manakala nilai outfit MNSQ berada di antara 0.45 – 2.54. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali item 140 (infit MNSQ = 1.70, outfit MNSQ = 2.54), item 156 (infit MNSQ = 2.06, outfit MNSQ = 1.46) dan item 160 (infit MNSQ = 1.46, outfit MNSQ = 1.61) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Manakala item 151 (infit MNSQ = 1.69), item 157 (infit MNSQ = 1.53), item 158 (infit MNSQ = 1.53) dan item 161 (infit MNSQ = 1.43) mempunyai nilai infit MNSQ yang melebihi 1.4. Item 147 (infit MNSQ = 0.51, outfit MNSQ = 0.51), item 164 (infit MNSQ = 0.45, outfit MNSQ = 0.54), item 144 (infit MNSQ = 0.50, outfit MNSQ = 0.45) dan item 146 (infit MNSQ = 0.38, outfit MNSQ = 0.47) mempunyai nilai infit dan outfit MNSQ yang menghampiri 0.6. Manakala item 138 (infit MNSQ = 0.54) mempunyai nilai infit MNSQ yang bawah 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.00 (item 140) dan 0.69 (item 144). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

*Domain Rohani 2.* Rajah 5.12 memetakan unsur-unsur bagi ketiga-tiga faset utama iaitu penilai, kanak-kanak dan item rohani bagi IPHKT 2 dengan menggunakan skala logit yang sepunya (*common log odds*). Bagi item-item rohani IPHKT 2 ini, item

150 adalah yang paling sukar. Lajur yang terakhir menunjukkan skala 5 mata yang digunakan dalam penskoran.

Logit	Kanak-kanak	Penilai	Item	Skala
6	*****	+	+	(5)
5	*	+	+	
	***			
4	+	+	+	
	**			
	***			
	****			
3	*****	+	+	
	**			
	*****			
	****			
	****			
2	***	+	+	
	*****			
	*****			---
	****		*	
	*****		*	
1	*****	+	***	
	*		*	4
	**	26 27	**	
		1 31	***	
	*	24 4 8	**	---
* 0	*	* 3 32 9	* *****	*
		20 29	***	3
		19	**	
		11 15 30	*	---
			*	2
-1	+	+	*	
			*	---
-2	+	+	+	
			*	
			*	
-3	+	+	+	(1)

Rajah 5.12. Pemetaan guru penilai, kanak-kanak dan item domain rohani IPHKT 2

Jadual 5.45 menunjukkan nilai penentukuran kesukaran logit (*difficulty calibrations*), ralat pengukuran, nilai infit dan outfit MNSQ serta nilai *point measure correlation* dan *point expected correlation* dan juga ringkasan statistik bagi faset item domain rohani IPHKT 2.

Jadual 5.45  
Laporan pengukuran item-item bagi domain rohani IPHKT 2

Item	Penentukuran Kesukaran Logit	S.E. Logit	Infit MNSQ	Outfit MNSQ	PtMea Corr	PtExp Corr
150	1.31	0.10	0.47	0.56	0.67	0.59
159	1.29	0.06	1.02	1.10	0.48	0.55
151	1.06	0.10	0.48	0.54	0.64	0.56
154	0.99	0.10	0.70	0.68	0.62	0.55
152	0.98	0.10	0.69	0.93	0.54	0.55
156	0.76	0.11	0.66	0.71	0.56	0.52
165	0.68	0.11	0.73	0.75	0.53	0.51
171	0.63	0.07	1.57	1.42	0.43	0.46
161	0.33	0.07	1.64	1.73	0.36	0.43
175	0.33	0.13	1.04	0.99	0.50	0.47
166	0.31	0.13	0.68	0.66	0.55	0.47
158	0.18	0.13	0.99	0.72	0.52	0.45
162	0.12	0.14	1.75	1.32	0.37	0.44
155	0.08	0.14	0.73	0.75	0.49	0.44
160	0.08	0.14	0.82	1.00	0.42	0.44
164	0.06	0.14	1.71	1.03	0.46	0.43
157	-0.02	0.14	1.10	0.64	0.50	0.42
169	-0.13	0.15	0.84	0.74	0.48	0.41
163	-0.27	0.16	1.04	1.47	0.37	0.39
174	-0.30	0.16	0.73	0.68	0.47	0.39
170	-0.35	0.16	0.89	0.70	0.45	0.38
172	-0.43	0.10	1.51	1.46	0.29	0.33
167	-0.89	0.12	1.21	0.99	0.30	0.28
176	-0.96	0.21	1.01	0.66	0.40	0.31
153	-1.22	0.13	1.14	1.14	0.22	0.25
173	-2.12	0.34	1.10	0.58	0.25	0.19
168	-2.54	0.42	1.25	0.68	0.19	0.16

Nota. Nilai khi-kuasdua : 1055.4; df: 26 ; Signifikan (kebarangkalian) : 0.00

Item 150 pada 1.31 logit (SE = 0.10) adalah item yang paling sukar, manakala item 168 pada -2.54 logit (SE = 0.42) adalah item yang paling mudah. Nilai khi-kuasdua adalah 1055.4, df adalah 26,  $p < 0.01$  bagi item menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara item-item. Ini mengesahkan bahawa terdapat variasi kesukaran yang signifikan di antara item-item.

Ketepatan pengukuran dilihat dari nilai ralat piawai. Nilai ralat piawai bagi domain rohani 2 berada di antara 0.06 hingga 0.42.

Jadual 5.45 menunjukkan julat nilai infit MNSQ berada di antara 0.47 – 1.75 manakala nilai outfit MNSQ berada di antara 0.54 – 1.73. Nilai dalam julat yang diterima iaitu di antara 0.6-1.4 (Linacre, 2012). Kesemua item-item didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch. Kecuali item 161 (infit MNSQ = 1.64, outfit MNSQ = 1.73), item 171 (infit MNSQ = 1.57, outfit MNSQ = 1.42) dan item 172 (infit MNSQ = 1.51, outfit MNSQ = 1.46) mempunyai nilai infit dan outfit MNSQ yang melebihi 1.4. Manakala item 162 (infit MNSQ = 1.75) dan item 164 (infit MNSQ = 1.71) mempunyai nilai infit MNSQ yang melebihi 1.4. Item 151 (infit MNSQ = 0.48, outfit MNSQ = 0.54) dan item 150 (infit MNSQ = 0.47, outfit MNSQ = 0.56) mempunyai nilai infit dan outfit MNSQ yang menghampiri 0.6. Manakala item 173 (outfit MNSQ = 0.58) mempunyai nilai outfit MNSQ yang menghampiri 0.6.

Nilai *point measure correlation* pula berada di antara nilai 0.19 (item 168) dan 0.67 (item 150). Semua nilai *point measure correlation* adalah positif yang menunjukkan ia bergerak pada satu arah. Ini mengesahkan lagi bahawa item-item adalah mengukur satu konstruk yang sama.

### **Menganalisis Dependability Score Kanak-kanak Menggunakan Teori-G**

***Apakah sumbangan setiap faset terhadap sumber varians mengikut Teori Generalizabiliti?***

*Domain Kognitif 1.* Jadual 5.46 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan kognitif. Analisis varian menunjukkan komponen varian item, (i) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 33.9%, diikuti dengan



komponen varian kanak-kanak sebanyak 24.8%. Seterusnya, sebanyak 24.2% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) sebanyak 17.1%.

Daripada analisis ini didapati komponen varian item (i) adalah paling tinggi ( $\sigma^2_i = 0.33447$  ; 33.9% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh item. Peratusan yang tinggi bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam ketiga-tiga item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak. Berdasarkan analisis ini juga, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.24533$ ; 24.8% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza.

Seterusnya, komponen varian menunjukkan bacaan yang besar adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.23933$ ; 24.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian penilai tersarang dalam kanak-kanak dan item, p:ki menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_{p:ki} = 0.16845$  ; 17.1% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai telah memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang sama semasa memberi skor untuk penilaian domain perkembangan kognitif. Berdasarkan kepada komponen varian

penilai yang rendah ini, boleh diandaikan sekiranya penilai lain yang telah mendapat latihan menilai prestasi kanak-kanak, tiada perbezaan skor di antara dua penilai tersebut.

Jadual 5.46  
Komponen Varians Faset Penilaian untuk Domain Perkembangan Kognitif IPHKT 1

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	865.13214		10.95104	0.24533	24.8
P:KI	1680	283.00000		0.16845	0.16845	17.1
Item	20	1083.23631		54.16182	0.33447	33.9
KI	1580	1022.43036		0.64711	0.23933	24.2
Jumlah	3359	3253.79881				100%

*Koeffisien G\_absolute* = 0.89

*Domain Kognitif 2.* Jadual 5.47 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan kognitif. Analisis varian menunjukkan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki), menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 38.3%, diikuti dengan komponen varian item (i) sebanyak 23.2%. Seterusnya, komponen varian kanak-kanak sebanyak 21.8% dan komponen varian terkecil ialah komponen varian interaksi di antara kanak-kanak dan item, (ki) sebanyak 17.1%.

Daripada analisis ini didapati komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) adalah paling tinggi ( $\sigma^2_{p:ki} = 0.18730$  ; 38.3% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh penilai. Peratusan yang tinggi bagi komponen penilai memberi makna penilai adalah berbeza dari segi aras

ketegasannya. Perbezaan aras ketegasan ini mempengaruhi skor yang diberikan kepada prestasi yang ditunjukkan oleh kanak-kanak.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari item adalah besar, ( $\sigma^2_i = 0.11379$ ; 23.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan item mempunyai aras kesukaran yang berbeza. Perbezaan kesukaran ini turut mempengaruhi skor prestasi kanak-kanak. Seterusnya, komponen varian menunjukkan bacaan yang besar adalah kanak-kanak (k) iaitu ( $\sigma_k = 0.10689$ ; 21.8% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan kanak-kanak adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Komponen varian terkecil ialah komponen varian interaksi di antara kanak-kanak dan item, ( $k_i$ ) menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_{ki} = 0.08159$ ; 16.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji.

Jadual 5.47  
*Komponen Varians Faset Penilaian untuk Domain Perkembangan Kognitif IPHKT 2*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	546.55925		7.19157	0.10689	24.8
P:KI	2464	461.50000		0.18730	0.18730	38.3
Item	31	554.11993		17.87484	0.11379	23.2
KI	2356	825.73945		0.35048	0.08159	16.7
Jumlah	4927	2387.91863				100%
<i>Koefisien G_absolute</i>	=	0.92				

*Domain Bahasa 1.* Jadual 5.48 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai

dalam penilaian berasas prestasi bagi domain perkembangan bahasa. Analisis varian menunjukkan komponen varian item, (i) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 40.7%, diikuti dengan komponen interaksi di antara kanak-kanak dan item, (ki) sebanyak 34.1%. Seterusnya, varian kanak-kanak sebanyak 15.5% dan komponen varian terkecil ialah komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) sebanyak 9.7%.

Daripada analisis ini didapati komponen varian item (i) adalah paling tinggi ( $\sigma^2_i = 0.72622$  ; 40.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh item. Peratusan yang tinggi bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam ketiga-tiga item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Berdasarkan analisis ini juga, komponen varian menunjukkan bacaan yang besar adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.60831$ ; 34.1 % daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Seterusnya, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.27710$ ; 15.5% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Komponen varian terkecil ialah komponen varian penilai tersarang dalam kanak-kanak dan item, p:ki menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_{p:ki} = 0.17361$ ; 9.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan di antara penilai dalam memberi skor kepada

kanak-kanak. Hal ini boleh disebabkan oleh penilai telah memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang sama semasa memberi skor untuk penilaian domain perkembangan bahasa. Berdasarkan kepada komponen varian penilai yang rendah ini, boleh diandaikan sekiranya penilai lain yang telah mendapat latihan menilai prestasi kanak-kanak, tiada perbezaan skor di antara dua penilai tersebut.

Jadual 5.48

*Komponen Varians Faset Penilaian untuk Domain Perkembangan Bahasa IPHKT 1*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	897.89444		11.36575	0.27710	15.5
P:KI	1440	250.00000		0.17361	0.17361	9.7
Item	17	1998.96250		117.58603	0.72622	40.7
KI	1343	1867.09306		1.39024	0.60831	34.1
Jumlah	3359	5013.95000				100%
<i>Koeffisien G_absolute</i>	=	0.78				

*Domain Bahasa 2.* Jadual 5.49 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan bahasa. Analisis varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) menunjukkan komponen varian menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 44.2%, diikuti dengan komponen varian kanak-kanak sebanyak 33.7%. Seterusnya, sebanyak 18.2% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian item sebanyak 3.9%.

Daripada analisis ini didapati komponen varian penilai tersarang dalam kanak-kanak dan item, (p:ki) menunjukkan nilai komponen varian yang paling tinggi, ( $\sigma^2_{p:ki} = 0.25373$ ; 44.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan

terdapat perbezaan di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai telah menunjukkan perbezaan semasa pemberian skor berdasarkan rubrik dan semua penilai mempunyai ketegasan yang berbeza semasa memberi skor untuk penilaian domain perkembangan bahasa.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.19388$ ; 33.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan kanak-kanak adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Seterusnya, komponen varian menunjukkan bacaan yang besar adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.10458$ ; 18.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian item (i) iaitu ( $\sigma_i = 0.02235$ ; 3.9% daripada jumlah keseluruhan komponen varians) Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini tidak dipengaruhi oleh item. Peratusan yang rendah bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah tidak berbeza dari segi aras kesukarannya. Aras kesukaran dalam item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Jadual 5.49

*Komponen Varians Penilaian untuk Domain Perkembangan Bahasa IPHKT 2*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	76	830.86436		10.93243	0.19388	33.7
P:KI	2079	527.50000		0.25373	0.25373	44.2
Item	26	101.53006		3.90500	0.02235	3.9
KI	1976	914.65512		0.46288	0.10458	18.2
Jumlah	4157	2374.54954				100%
Koeffisien <i>G_absolute</i>	=	0.95				

*Domain Fizikal 1.* Jadual 5.50 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan fizikal. Analisis varian menunjukkan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 65%, diikuti dengan komponen varian kanak-kanak sebanyak 34%. Seterusnya, sebanyak 1.0% ialah varian item (i), dan komponen varian terkecil ialah, interaksi di antara kanak-kanak dan item, (ki) sebanyak 0.0%.

Daripada analisis ini didapati komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) adalah paling tinggi ( $\sigma^2_{p:ki} = 0.61958$ ; 65.0% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh penilai. Peratusan yang tinggi bagi komponen penilai memberi makna penilai dalam penilaian ini adalah berbeza dari segi aras ketegasannya. Perbezaan aras ketegasan dalam kalangan penilai ini mempengaruhi skor prestasi yang diberikan kepada kanak-kanak.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.32435$ ; 34% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Seterusnya, komponen varian item menunjukkan bacaan yang rendah iaitu adalah iaitu ( $\sigma^2_i = 0.0092$ ; 1.0% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang ketara aras kesukaran item-item yang diuji. Komponen varian terkecil ialah komponen varian interaksi di antara kanak-kanak dan item (ki) menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_{ki} = -0.07004$ ; 0.0% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tiada perbezaan di antara kanak-kanak terhadap item-item yang diuji.

Hal ini boleh disebabkan oleh kanak-kanak boleh memahami item dan melakukannya dan semua kanak-kanak mempunyai konsistensi yang sama terhadap item untuk penilaian domain perkembangan fizikal oleh penilai. Berdasarkan kepada komponen varian komponen kanak-kanak dan item yang rendah ini, boleh diandaikan sekiranya kanak-kanak lain diberi item yang sama untuk diuji, tiada perbezaan interaksi antara kanak-kanak dengan item tersebut.

Jadual 5.50

*Komponen Varians Penilaian untuk Domain Perkembangan Fizikal IPHKT 1*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	806.59958		10.21012	0.32435	34.0
P:KI	1200	743.50000		0.61958	0.61958	65.0
Item	14	27.46833		1.96202	0.00927	1.0
KI	1106	530.33167		0.47950	-0.07004	0.0
Jumlah	2399	2107.89958				100%
<i>Koeffisien G_absolute</i>	=	0.94				

*Domain Fizikal 2.* Jadual 5.51 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan fizikal. Analisis varian menunjukkan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 33.1%, diikuti dengan komponen varian ialah interaksi di antara kanak-kanak dan item, (ki) sebanyak 32.9%. Seterusnya, komponen varian item (i) sebanyak 18.0 % dan komponen varian terkecil ialah komponen varian kanak-kanak (k) sebanyak 15.9%.

Daripada analisis ini didapati komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) adalah paling tinggi ( $\sigma^2_{p:ki} = 0.12208$ ; 33.1% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan



di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang berbeza semasa memberi skor untuk penilaian domain perkembangan fizikal. Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh penilai.

Berdasarkan analisis ini juga, komponen varian menunjukkan bacaan yang besar adalah interaksi di antara kanak-kanak dan item ( $\sigma_{ki} = 0.12152$ ; 32.9% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Seterusnya, menunjukkan komponen varian dari item, (i) menunjukkan bacaan sederhana, ( $\sigma^2_i = 0.06642$ ; 18.0% daripada jumlah keseluruhan komponen varians). Peratusan yang sederhana bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam kesemua item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak secara sederhana. Komponen varian terkecil ialah komponen varian kanak-kanak (k) menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_k = 0.05879$  ;15.9% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, terdapat sedikit perbezaan dalam kebolehan kanak-kanak, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza yang sederhana.

Jadual 5.51

*Komponen Varians Faset Penilaian untuk Domain Perkembangan Fizikal IPHKT 2*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	76	251.13766		3.30444	0.05879	15.9
P:KI	1925	235.00000		0.12208	0.12208	33.1
Item	24	254.23377		10.59307	0.06642	18.0
KI	1824	665.96623		0.36511	0.12152	32.9
Jumlah	3849	1406.33766				100%
<i>Koeffisien G_absolute</i>	=	0.86				

*Domain Kreativiti 1.* Jadual 5.52 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan kreativiti. Analisis varian menunjukkan komponen varian interaksi di antara kanak-kanak dan item, (ki) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak 32.1%, diikuti dengan komponen kanak-kanak, (k) sebanyak 31.4%. Seterusnya, sebanyak 29% ialah penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) dan komponen varian terkecil ialah komponen varian item sebanyak 7.5%.

Daripada analisis ini didapati komponen varian interaksi di antara kanak-kanak dan item (ki) adalah paling tinggi ( $\sigma^2_{ki} = 0.21213$ ; 32.1% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh item. Peratusan yang tinggi bagi komponen item memberi makna Ini menunjukkan terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.20783$ ; 31.4% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan

kanak-kanak adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Seterusnya, komponen varian menunjukkan bacaan yang besar adalah penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) iaitu ( $\sigma_{p:ki} = 0.19156$ ; 29% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat sedikit perbezaan di antara penilai dalam memberi skor terhadap respon kanak-kanak kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian item (i), menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_i = 0.04983$ ; 7.5% daripada jumlah keseluruhan komponen varians). Ini menunjukkan item mudah difahami oleh kanak-kanak. Hal ini boleh disebabkan kanak-kanak memahami item dan mempunyai konsistensi yang sama semasa memberi respon untuk penilaian domain perkembangan kreativiti. Berdasarkan kepada komponen varian item yang rendah ini, boleh diandaikan sekiranya kanak-kanak lain diuji dengan item yang sama, maka tiada perbezaan yang banyak terhadap respon yang diterima.

Jadual 5.52

*Komponen Varians Faset Penilaian untuk Domain Perkembangan Kreativiti IPHKT 1*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	705.40469		8.92917	0.20783	31.4
P:KI	1600	306.50000		0.19156	0.19156	29.0
Item	19	163.18594		8.58873	0.04983	7.5
KI	1501	924.33906		0.61582	0.21213	32.1
Jumlah	3199	2099.42969				100%
Koefisien $G_{absolute}$	=	0.92				

*Domain Kreativiti 2.* Jadual 5.53 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan kreativiti. Analisis varian menunjukkan komponen varian item, (i) menyumbang peratusan yang tertinggi

daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 36.0%, diikuti dengan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) sebanyak 23.2%. Seterusnya, sebanyak 20.7% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian kanak-kanak (k) sebanyak 20.2%.

Daripada analisis ini didapati komponen varian item (i) adalah paling tinggi ( $\sigma^2_i = 0.22488$ ; 36.0% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi oleh item. Peratusan yang tinggi bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam ketiga-tiga item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Berdasarkan analisis ini juga, komponen varian penilai tersarang dalam kanak-kanak dan item, p:ki menunjukkan nilai komponen varian yang tinggi, ( $\sigma^2_{p:ki} = 0.14461$ ; 23.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai telah memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang tidak sama semasa memberi skor untuk penilaian domain perkembangan kreativiti. Seterusnya, komponen varian menunjukkan bacaan yang besar adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.12914$ ; 20.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian dari kanak-kanak (k), ( $\sigma^2_k = 0.12588$ ; 20.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, tiada perbezaan kebolehan

kanak-kanak, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang hampir sama.

Jadual 5.53

*Komponen Varians Faset Penilaian untuk Domain Perkembangan Kreativiti IPHKT 2*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	76	528.10090		6.94870	0.12588	20.2
P:KI	2002	289.50000		0.14461	0.14461	23.2
Item	25	875.85639		35.03426	0.22488	36.0
KI	1900	765.47053		0.40288	0.12914	20.7
Jumlah	4003	2458.92782				100%
<i>Koeffisien G_absolute</i>	=	0.88				

*Domain Sosioemosi 1.* Jadual 5.54 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan sosioemosi. Analisis varian menunjukkan komponen varian item menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 47.9%, diikuti penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) sebanyak 18.6%. Seterusnya, sebanyak 18.2% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian kanak-kanak sebanyak 15.3%.

Daripada analisis ini didapati komponen varian item (i), menunjukkan nilai komponen varian yang paling tinggi, ( $\sigma_{ki} = 0.16381$ ; 47.9% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini dipengaruhi oleh item. Peratusan yang tinggi bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kefahamannya. Perbezaan aras kefahaman dalam item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Berdasarkan analisis ini juga, komponen varian penilai tersarang dalam kanak-kanak dan item, ( $\sigma^2_{p:ki}$ ) menunjukkan nilai komponen varian yang sederhana, ( $\sigma^2_{p:ki} = 0.00713$  ; 18.6% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang sederhana di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai telah memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang sama semasa memberi skor untuk penilaian domain perkembangan sosioemosi. Seterusnya, komponen varian menunjukkan bacaan yang sederhana juga adalah interaksi di antara kanak-kanak dan item ( $\sigma^2_{ki}$ ) iaitu ( $\sigma^2_{ki} = 0.01130$ ; 18.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang sederhana di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian daripada kanak-kanak iaitu ( $\sigma^2_k = 0.04062$ ; 15.3% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah kecil, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang tertentu dan hampir sama.

Jadual 5.54

*Komponen Varians Faset Penilaian Domain Perkembangan Sosioemosi IPHKT 1*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	1838.13278		23.26750	0.24859	15.3
P:KI	3600	1090.00000		0.30278	0.30278	18.6
Item	44	5530.69111		125.69753	0.78002	47.9
KI	3476	3108.64222		0.89432	0.29577	18.2
Jumlah	7199	11567.46611				100%
<i>Koeffisien G absolute</i>	=	0.90				

*Domain Sosioemosi 2.* Jadual 5.55 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan sosioemosi. Analisis varian menunjukkan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki), menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 42.2%, diikuti dengan komponen varian kanak-kanak sebanyak 33%. Seterusnya, sebanyak 17.7% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian item sebanyak 7.2%.

Daripada analisis ini didapati komponen varian penilai tersarang dalam kanak-kanak dan item, (p:ki) menunjukkan nilai komponen varian yang paling tinggi, ( $\sigma^2_{p:ki} = 0.24942$  ; 42.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan di antara penilai dalam memberi skor kepada kanak-kanak. Hal ini boleh disebabkan oleh penilai telah memahami pemberian skor berdasarkan rubrik dan semua penilai mempunyai konsistensi yang tidak sama semasa memberi skor untuk penilaian domain perkembangan sosioemosi.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari kanak-kanak adalah besar, ( $\sigma^2_k = 0.19502$ ; 33.0% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Seterusnya, komponen varian menunjukkan bacaan yang sederhana adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.10476$ ; 17.7% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan yang sederhana di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian item (i), ( $\sigma_{ki}=0.04232$ ; 7.2% daripada jumlah keseluruhan komponen varians). Ini

menunjukkan *dependability score* kanak-kanak dalam ujian ini tidak dipengaruhi oleh item. Peratusan yang rendah bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam ketiga-tiga item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Jadual 5.55

*Komponen Varians Faset Penilaian untuk Domain Perkembangan Sosioemosi IPHKT*

2

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	76	1190.97369		15.67071	0.19502	33.0
P:KI	3003	749.00000		0.24942	0.24942	42.2
Item	38	265.08425		6.97590	0.04232	7.2
KI	2888	1325.40293		0.10476	0.10476	17.7
Jumlah	6005	3530.46087				100%
<i>Koeffisien G_absolute</i>	=	0.97				

*Domain Rohani 1.* Jadual 5.56 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan rohani. Analisis varian menunjukkan komponen varian item, (i) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 63.2%, diikuti dengan komponen varian interaksi di antara kanak-kanak dengan item (ki) sebanyak 17.3%. Seterusnya, penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki) sebanyak 10.7%. Komponen varian terkecil ialah komponen varian kanak-kanak sebanyak 8.8%.

Daripada analisis ini didapati komponen varian item (i) adalah paling tinggi ( $\sigma^2_i = 1.19747$ ; 63.2% daripada jumlah keseluruhan komponen varians). Ini menunjukkan *dependability score* kanak-kanak dalam ujian ini banyak dipengaruhi



oleh item. Peratusan yang tinggi bagi komponen item memberi makna item yang diuji dalam penilaian ini adalah berbeza dari segi aras kesukarannya. Perbezaan aras kesukaran dalam ketiga-tiga item ini mempengaruhi prestasi yang ditunjukkan oleh kanak-kanak.

Berdasarkan analisis ini juga, menunjukkan komponen varian dari interaksi di antara kanak-kanak-item adalah lebih kecil, ( $\sigma^2_{ki} = 0.32714$ ; 17.3% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah kecil, ini bermakna kanak-kanak yang menjadi responden kajian tidak mempunyai kebolehan yang berbeza. Seterusnya, komponen varian menunjukkan bacaan yang kecil juga adalah komponen varian penilai tersarang dalam kanak-kanak dan item iaitu ( $\sigma_{p:ki} = 0.20331$ ; 10.7% daripada jumlah keseluruhan komponen varians). Bacaan yang rendah ini menunjukkan tidak terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen kanak-kanak (k), menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_k = 0.16682$  ; 8.8% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan di antara kanak-kanak dalam memberi respon semasa penilaian. Hal ini boleh disebabkan oleh kanak-kanak telah memahami atau tidak memahami untuk penilaian domain perkembangan rohani. Berdasarkan kepada komponen varian kanak-kanak yang rendah ini, boleh diandaikan bahawa pada peringkat umur ini, kanak-kanak perlu pemahaman lebih jelas kepada aspek rohani yang ingin dinilai. Maka, kefahaman yang tidak dicapai atau sebaliknya menjadikan tiada perbezaan antara kanak-kanak semasa proses penilaian dijalankan.

Jadual 5.56

*Komponen Varians Penilaian untuk Domain Perkembangan Rohani IPHKT 1*

Faset	Df	Jumlah Dua	Kuasa	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	79	963.91176		12.20141	0.16682	8.8
P:KI	2720	553.00000		0.20331	0.20331	10.7
Item	33	6350.95074		192.45305	1.19747	63.2
KI	2607	2235.72574		0.85759	0.32714	17.3
Jumlah	5439	10103.58824				100%
<i>Koeffisien G_absolute</i>	=	0.78				

*Domain Rohani 2.* Jadual 5.57 menunjukkan komponen varian setiap faset yang menyumbang kepada perbezaan skor kanak-kanak yang dinilai oleh penilai dalam penilaian berasas prestasi bagi domain perkembangan rohani. Analisis varian menunjukkan komponen varian penilai yang tersarang (*nested*) dalam kanak-kanak dan item, (p:ki, (k) menyumbang peratusan yang tertinggi daripada jumlah keseluruhan varian, iaitu sebanyak sebanyak 31.5%, diikuti dengan komponen varian kanak-kanak sebanyak 29.4%. Seterusnya, sebanyak 25.8% ialah interaksi di antara kanak-kanak dan item, (ki) dan komponen varian terkecil ialah komponen varian item sebanyak 13.3%.

Daripada analisis ini didapati, komponen varian penilai tersarang dalam kanak-kanak dan item, p:ki menunjukkan nilai komponen varian yang tinggi, ( $\sigma^2_{p:ki} = 0.17292$ ; 31.5% daripada jumlah keseluruhan komponen varians). Ini menunjukkan daripada responden kanak-kanak ini, perbezaan kebolehan pelajar adalah besar, ini bermakna kanak-kanak yang menjadi responden kajian mempunyai kebolehan yang berbeza. Komponen varian kanak-kanak (k) menunjukkan bacaan yang kedua tinggi iaitu ( $\sigma^2_{p:ki} = 0.02721$ ; 29.4% daripada jumlah keseluruhan komponen varians). Ini menunjukkan terdapat perbezaan kebolehan kanak-kanak yang ditunjukkan terhadap respon kebolehan kanak-kanak.

Seterusnya, komponen varian menunjukkan bacaan yang sederhana adalah interaksi di antara kanak-kanak dan item (ki) iaitu ( $\sigma_{ki} = 0.00774$ ; 25.8% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan yang ketara di antara kanak-kanak dalam memberi respon kepada item-item yang diuji. Komponen varian terkecil ialah komponen varian item, (i) menunjukkan nilai komponen varian yang rendah, ( $\sigma^2_{p:ki} = 0.02035$ ; 13.3% daripada jumlah keseluruhan komponen varians). Ini menunjukkan tidak terdapat perbezaan di antara kanak-kanak dalam memberi respon semasa penilaian. Hal ini boleh disebabkan oleh kanak-kanak telah memahami atau tidak memahami untuk penilaian domain perkembangan rohani. Berdasarkan kepada komponen varian kanak-kanak yang rendah ini, boleh diandaikan bahawa pada peringkat umur ini, kanak-kanak perlu pemahaman lebih jelas kepada aspek rohani yang ingin dinilai. Maka kefahaman yang tidak dicapai atau sebaliknya menjadikan tiada perbezaan antara kanak-kanak semasa proses penilaian dijalankan.

Jadual 5.57

*Komponen Varians Penilaian untuk Domain Perkembangan Rohani IPHKT 2*

Faset	Df	Jumlah Kuasa Dua	Min Kuasa Dua	Komponen Varians	% Varians
Kanak-kanak, K	76	697.38672	9.17614	0.16148	29.4
P:KI	2079	359.50000	0.17292	0.17292	31.5
Item	26	304.92689	11.72796	0.07319	13.3
KI	1976	901.92496	0.45644	0.14176	25.8
Jumlah	4157	2263.73858			100%
<i>Koeffisien G_absolute</i>	=	0.94			

**Apakah *dependability score koeffisien G* bagi faset domain mengikut Teori Generalizabiliti?**

*Domain Kognitif 1. Koeffisien relative (0.94) dan Koeffisien absolute (0.89) menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk*

kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.58, sebanyak 64% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 64% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 36% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 64% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 36% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.17702. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.49531 untuk serakan skor sebenar (*true score*).

Jadual 5.58  
*Dependability Score Koeffisien Domain Kognitif IPHKT 1*

<i>Source of variance</i>	<i>Differentiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.24533		.....		.....	
	.....	P :KI	0.00401	26.0	0.00401	12.8
	.....	I	.....		0.01593	50.8
	.....	KI	0.01140	74.0	0.01140	36.4
<i>Sum of variances</i>	0.24533		0.01541	100%	0.03133	100%
<i>Standard deviation</i>	0.49531		Relative SE: 0.12413		Absolute SE: 0.17702	
<i>Coef_G relative</i>	0.94					
<i>Coef_G absolute</i>	0.89					

*Domain Kognitif 2. Koeffisien relative (0.95) dan Koeffisien absolute (0.92)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.59, sebanyak 85% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 85% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 15% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 85% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 15% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.09504. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.32694 untuk serakan skor sebenar (*true score*).

Jadual 5.59  
*Dependability Score Koeffisien Domain Kognitif IPHKT 2*

<i>Source of variance</i>	<i>Differentiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.10689		.....		.....	
	.....	P :KI	0.00293	53.4	0.00293	32.4
	.....	I	.....		0.00356	39.4
	.....	KI	0.00255	46.6	0.00255	28.2
<i>Sum of variances</i>	0.10689		0.00548	100%	0.03133	100%
<i>Standard deviation</i>	0.32694		<i>Relative SE: 0.07400</i>		<i>Absolute SE: 0.09504</i>	
<i>Coef_G relative</i>	0.95					
<i>Coef_G absolute</i>	0.92					

*Domain Bahasa 1. Koeffisien relative (0.88) dan Koeffisien absolute (0.78)* menunjukkan nilai di bawah dan menghampiri nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian masih baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai koeffisien G sedikit di bawah nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis Jadual 5.60, sebanyak 61% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 61% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 39% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 61% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 39% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.28100. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.52640 untuk serakan skor sebenar (*true score*).

Jadual 5.60  
*Dependability Score Koeffisien Domain Bahasa IPHKT 1*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.27710		.....		.....	
	.....	P :KI	0.00482	12.5	0.00482	6.1
	.....	I	.....		0.04035	51.1
	.....	KI	0.03380	87.5	0.03380	42.8
<i>Sum of variances</i>	0.27710		0.03862	100%	0.07896	100%
<i>Standard deviation</i>	0.52640		<i>Relative SE: 0.19651</i>		<i>Absolute SE: 0.28100</i>	
<i>Coef_G relative</i>	0.88					
<i>Coef_G absolute</i>	0.78					

*Domain Bahasa 2. Koeffisien relative (0.96) dan Koeffisien absolute (0.95)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.61, sebanyak 90% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 90% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 10% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 90% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 10% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.09695. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.44032 untuk serakan skor sebenar (*true score*).

Jadual 5.61  
*Dependability Score Koeffisien Domain Bahasa IPHKT 2*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.19388		.....		.....	
	.....	P :KI	0.00470	54.8	0.00470	50.0
	.....	I	.....		0.00083	8.8
	.....	KI	0.00387	45.2	0.00387	41.2
<i>Sum of variances</i>	0.19388		0.00857	100%	0.00940	100%
<i>Standard deviation</i>	0.44032		<i>Relative SE: 0.09258</i>		<i>Absolute SE: 0.09695</i>	
<i>Coef_G relative</i>	0.96					
<i>Coef_G absolute</i>	0.95					

*Domain Fizikal 1. Koeffisien relative (0.94) dan Koeffisien absolute (0.94)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.62, sebanyak 88% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 88% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 12% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 88% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 12% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.14584. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.56952 untuk serakan skor sebenar (*true score*).



Jadual 5.62

*Dependability Score Koeffisien Domain Fizikal IPHKT 1*

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
K	0.32435		.....		.....	
	.....	P :KI	0.02065	100.0	0.02065	97.1
	.....	I	.....		0.00062	2.9
	.....	KI	(0.00000)	0.0	(0.00000)	0.0
Sum of variances	0.32435		0.02065	100%	0.02127	100%
Standard deviation	0.56952		Relative SE: 0.14371		Absolute SE: 0.14584	
Coef_G relative	0.94					
Coef_G absolute	0.94					

*Domain Fizikal 2. Koeffisien relative (0.89) dan Koeffisien absolute (0.86)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.63, sebanyak 74% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 74% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 26% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 74% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 26% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.09979. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.24246 untuk serakan skor sebenar (*true score*).

Jadual 5.63  
*Dependability Score Koeffisien Domain Fizikal IPHKT 2*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.05879	.....	.....	.....	.....	.....
	.....	P :KI	0.00244	33.4	0.00244	24.5
	.....	I	.....	.....	0.00266	26.7
	.....	KI	0.00486	66.6	0.00486	48.8
<i>Sum of variances</i>	0.05879		0.01541	100%	0.03133	100%
<i>Standard deviation</i>	0.24246		<i>Relative SE: 0.08545</i>		<i>Absolute SE: 0.09979</i>	
<i>Coef_G relative</i>	0.89					
<i>Coef_G absolute</i>	0.86					

*Domain Kreativiti 1. Koeffisien relative (0.93) dan Koeffisien absolute (0.92)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.64, sebanyak 85% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 85% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 15% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 85% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 15% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.13374. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.49531 untuk serakan skor sebenar (*true score*).

Jadual 5.64

*Dependability Score Koeffisien Domain Kreativiti IPHKT 1*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.20783	.....	.....	.....	.....	.....
	.....	P :KI	0.00479	31.1	0.00479	26.8
	.....	I	.....	.....	0.00249	13.9
	.....	KI	0.01061	68.9	0.01061	59.3
<i>Sum of variances</i>	0.24533		0.01541	100%	0.03133	100%
<i>Standard deviation</i>	0.49531		<i>Relative SE: 0.12408</i>		<i>Absolute SE: 0.13374</i>	
<i>Coef_G relative</i>	0.93					
<i>Coef_G absolute</i>	0.92					

*Domain Kreativiti 2.* Koeffisien *G relative* (0.94) dan koeffisien *G absolute* (0.88) menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai koeffisien *G* di atas nilai konvensional. Koeffisien *G absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.65, sebanyak 77% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 77% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 23% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 77% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 23% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.12805. *Standard error*

menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.49531 untuk serakan skor sebenar (*true score*).

Jadual 5.65

*Dependability Score Koeffisien Domain Kreativiti IPHKT 2*

<i>Source of variance</i>	<i>Differentiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.12588		.....		.....	
	.....	P :KI	0.00278	35.9	0.00278	17.0
	.....	I	.....		0.00865	52.7
	.....	KI	0.00497	64.1	0.00497	30.3
<i>Sum of variances</i>	0.24533		0.00775	100%	0.01640	100%
<i>Standard deviation</i>	0.49531		<i>Relative SE: 0.08802</i>		<i>Absolute SE: 0.12805</i>	
<i>Coef_G relative</i>	0.94					
<i>Coef_G absolute</i>	0.88					

*Domain Sosioemosi 1.* Koeffisien *G relative* (0.96) dan koeffisien *G absolute* (0.90) menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai koeffisien *G* di atas nilai konvensional. Koeffisien *G absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.66, sebanyak 81% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 81% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 19% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 81% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 19% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-

kanak adalah kecil di mana *standard error absolute* adalah 0.16514. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.49859 untuk serakan skor sebenar (*true score*).

Jadual 5.66

*Dependability Score Koeffisien Domain Sosioemosi IPHKT 1*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.24859	.....	.....	.....	.....	.....
	.....	P :KI	0.00336	33.9	0.00336	12.3
	.....	I	.....	.....	0.01733	63.6
	.....	KI	0.00657	66.1	0.00657	24.1
<i>Sum of variances</i>	0.24859		0.00994	100%	0.00697	100%
<i>Standard deviation</i>	0.49859		<i>Relative SE: 0.09968</i>		<i>Absolute SE: 0.16514</i>	
<i>Coef_G relative</i>	0.96					
<i>Coef_G absolute</i>	0.90					

*Domain Sosioemosi 2. Koeffisien relative (0.97) dan Koeffisien absolute (0.97)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.67, sebanyak 94% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 94% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 6% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 94% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 6% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-

kanak adalah kecil di mana *standard error absolute* adalah 0.08348. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.44161 untuk serakan skor sebenar (*true score*).

Jadual 5.67

*Dependability Score Koeffisien Domain Sosioemosi IPHKT 2*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.19502		.....		.....	
	.....	P :KI	0.00320	54.3	0.00320	45.9
	.....	I	.....		0.00109	15.6
	.....	KI	0.00269	45.7	0.00269	38.5
<i>Sum of variances</i>	0.19502		0.00588	100%	0.00697	100%
<i>Standard deviation</i>	0.44161		<i>Relative SE: 0.07671</i>		<i>Absolute SE: 0.08348</i>	
<i>Coef_G relative</i>	0.97					
<i>Coef_G absolute</i>	0.97					

*Domain Rohani 1. Koeffisien G relative* (0.93) dan *koeffisien G absolute* (0.78) menunjukkan nilai menghmpiri nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian dianggap baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai koeffisien G menghampiri nilai konvensional. Koeffisien *G absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.68, sebanyak 61% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 61% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 39% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 61% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 39% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga

menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.21870. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.40844 untuk serakan skor sebenar (*true score*).

Jadual 5.68  
*Dependability Score Koeffisien Domain Rohani IPHKT 1*

<i>Source of variance</i>	<i>Differ-entiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.16682		.....		.....	
	.....	P :KI	0.00299	23.7	0.00299	6.3
	.....	I	.....		0.03522	73.6
	.....	KI	0.00962	76.3	0.00962	20.1
<i>Sum of variances</i>	0.16682		0.01261	100%	0.03133	100%
<i>Standard deviation</i>	0.40844		<i>Relative SE: 0.11230</i>		<i>Absolute SE: 0.21870</i>	
<i>Coef_G relative</i>	0.93					
<i>Coef_G absolute</i>	0.78					

*Domain Rohani 2. Koeffisien G relative (0.95) dan Koeffisien G absolute (0.94)* menunjukkan nilai di atas nilai konvensional yang diterima iaitu 0.8. Reka bentuk kajian sangat baik untuk menganalisis kebergantungan skor kanak-kanak kerana nilai *Koeffisien* di atas nilai konvensional. *Koeffisien absolute* dipertimbangkan kerana kajian ini adalah untuk menilai *dependability score* kanak-kanak secara individu berdasarkan sumbangan komponen varian dalam penilai yang berbeza. Daripada analisis dalam Jadual 5.69, sebanyak 88% dapatan daripada skor kanak-kanak adalah bergantung (*attributable*) kepada skor universe (*universe score*). Ini bermakna 88% daripada keseluruhan kajian ini dapat dijelaskan. Manakala 12% sahaja skor dapatan bergantung (*attributable*) kepada kesan rawak yang tidak dapat dikenalpasti.

Rekabentuk ini menghasilkan kebolehpercayaan ukuran atau ukuran yang diyakini. Ianya juga boleh diinterpretasikan sebagai 88% faktor yang menyumbang kepada varians skor kanak-kanak dapat dijelaskan manakala 12% faktor penyumbang didapati daripada sumber ralat yang tidak dapat dikenalpasti. Dapatan kajian juga

menunjukkan ralat standard (*standard error*) berkaitan dengan keputusan skor kanak-kanak adalah kecil di mana *standard error absolute* adalah 0.10566. *Standard error* menunjukkan nilai lebih kecil daripada *estimated standard deviation* 0.40184 untuk serakan skor sebenar (*true score*).

Jadual 5.69  
*Dependability Score Koeffisien Domain Rohani IPHKT 2*

<i>Source of variance</i>	<i>Differentiation variance</i>	<i>Source of variance</i>	<i>Relative error variance</i>	<i>% relative</i>	<i>Absolute error variance</i>	<i>% absolute</i>
K	0.16148	.....	.....	.....	.....	.....
	.....	P :KI	0.00320	37.9	0.00320	28.7
	.....	I	.....	.....	0.00271	24.3
	.....	KI	0.00525	62.1	0.00525	47.0
<i>Sum of variances</i>	0.16148		0.00845	100%	0.01116	100%
<i>Standard deviation</i>	0.40184		<i>Relative SE: 0.09194</i>		<i>Absolute SE: 0.10566</i>	
<i>Coef_G relative</i>	0.95					
<i>Coef_G absolute</i>	0.94					

**Apakah nilai *optimization* yang terbaik bagi faset kanak-kanak dan penilai menggunakan kajian D untuk meningkatkan *Koeffisien*?**

*Domain Kognitif 1*. Dalam kajian–G, *Coef\_G relative* ( $\hat{\epsilon}p^2$ ) memaparkan tahap perbezaan ralat *relative* varian (*relative error variance*). Dalam kajian–D *Coef\_G absolute* phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.70 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.70  
*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kognitif IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
<i>Coef_G relative</i> ( $\hat{\epsilon}p^2$ )	0.94	0.95	0.95	0.95	0.94	0.94
<i>Coef_G absolute</i> ( $\Phi$ )	0.89	0.89	0.89	0.89	0.89	0.89



Dalam kajian ini, *Coef\_G absolute phi* ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan kognitif dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.70, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, *Coef\_G absolute phi* ( $\Phi$ ) kekal pada 0.89, iaitu satu nilai yang tinggi dan diterima. Nilai *Coef\_G absolute phi* ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan kognitif dikekalkan untuk nilai *coef\_G absolute phi* ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Kognitif 2.* Dalam kajian-G, *Coef\_G relative* ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute phi* ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.71 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.71

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kognitif IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.95	0.96	0.96	0.96	0.95	0.95
Coef_G absolute ( $\Phi$ )	0.92	0.93	0.93	0.93	0.92	0.92

Dalam kajian ini, Coef\_G absolute phi ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan kognitif dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.71, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambil kira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G absolute phi ( $\Phi$ ) kekal pada 0.92, iaitu satu nilai yang tinggi dan diterima. Nilai Coef\_G absolute phi ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan kognitif dikekalkan untuk nilai coef\_G absolute phi ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Bahasa 1.* Dalam kajian-G, Coef\_G relative ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat relative varian (*relative error variance*). Dalam kajian-D Coef\_G absolute phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute*

*error variance*). Jadual 5.72 menunjukkan perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.72

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Bahasa IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.88	0.88	0.88	0.88	0.88	0.88
Coef_G absolute ( $\Phi$ )	0.78	0.78	0.78	0.78	0.78	0.78

Dalam kajian ini, Coef\_G *absolute* phi ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan bahasa dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.72, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambil kira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G *absolute* phi ( $\Phi$ ) kekal pada 0.78, iaitu satu nilai yang masih boleh diterima. Nilai Coef\_G *absolute* phi ( $\Phi$ ) ini menghampiri nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan bahasa dikekalkan untuk nilai *coef\_G absolute* phi ( $\Phi$ ) kajian ini yang masih diterima dan mempunyai nilai kebolehpercayaan yang tinggi.

*Domain Bahasa 2.* Dalam kajian–G, Coef\_G relative ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian–D Coef\_G *absolute phi* ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.73 menunjukkan perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.73

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Bahasa IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.96	0.97	0.97	0.97	0.96	0.96
Coef_G <i>absolute phi</i> ( $\Phi$ )	0.95	0.96	0.96	0.96	0.95	0.95

Dalam kajian ini, Coef\_G *absolute phi* ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan bahasa dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.73, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G *absolute phi* ( $\Phi$ ) kekal pada 0.95, iaitu satu nilai yang tinggi dan diterima. Nilai Coef\_G *absolute phi* ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas

prestasi kepada domain perkembangan bahasa dikekalkan untuk nilai *coef\_G absolute phi* ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Fizikal 1.* Dalam kajian-G, *Coef\_G relative* ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute phi* ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.74 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.74

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Fizikal IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
<i>Coef_G relative</i> ( $\hat{E}p^2$ )	0.94	0.97	0.97	0.96	0.94	0.95
<i>Coef_G absolute phi</i> ( $\Phi$ )	0.94	0.97	0.97	0.96	0.94	0.95

Dalam kajian ini, *Coef\_G absolute phi* ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan fizikal dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.74, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambil kira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, *Coef\_G absolute phi* ( $\Phi$ ) kekal pada 0.89, iaitu satu nilai yang tinggi dan diterima. Nilai *Coef\_G absolute phi* ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa,

kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan fizikal dikekalkan untuk nilai *coef\_G absolute* phi ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Fizikal 2.* Dalam kajian-G, *Coef\_G relative* ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute* phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak *varian ralat* (*absolute error variance*). Jadual 5.75 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.75

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Fizikal IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
<i>Coef_G relative</i> ( $\hat{E}p^2$ )	0.89	0.91	0.91	0.90	0.89	0.89
<i>Coef_G absolute</i> ( $\Phi$ )	0.86	0.87	0.87	0.86	0.85	0.85

Dalam kajian ini, *Coef\_G absolute* phi ( $\Phi$ ) sahaja diambilkira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan fizikal dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.75, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, *Coef\_G absolute* phi ( $\Phi$ ) kekal pada 0.86, iaitu satu nilai yang tinggi dan diterima. Nilai *Coef\_G absolute* phi ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-

lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan fizikal dikekalkan untuk nilai *coef\_G absolute* phi ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Kreativiti 1.* Dalam kajian-G, *Coef\_G relative* ( $\hat{\epsilon}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute* phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.76 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.76

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kreativiti IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
<i>Coef_G relative</i> ( $\hat{\epsilon}p^2$ )	0.93	0.94	0.94	0.94	0.93	0.93
<i>Coef_G absolute</i> ( $\Phi$ )	0.92	0.93	0.93	0.93	0.92	0.92

Dalam kajian ini, *Coef\_G absolute* phi ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan kreativiti dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.76, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, *Coef\_G absolute* phi ( $\Phi$ ) kekal pada 0.92, iaitu satu nilai yang tinggi dan diterima. Nilai *Coef\_G absolute* phi ( $\Phi$ ) ini melebihi nilai konvensional yang

diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan kreativiti dikekalkan untuk nilai *coef\_G absolute*  $\phi$  ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Kreativiti 2.* Dalam kajian-G, *Coef\_G relative* ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute*  $\phi$  ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.77 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.77

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Kreativiti IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
<i>Coef_G relative</i> ( $\hat{E}p^2$ )	0.94	0.95	0.95	0.95	0.94	0.94
<i>Coef_G absolute</i> ( $\Phi$ )	0.88	0.89	0.89	0.89	0.88	0.88

Dalam kajian ini, *Coef\_G absolute*  $\phi$  ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan kreativiti dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.77, didapati bilangan



kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G *absolute* phi ( $\Phi$ ) kekal pada 0.88, iaitu satu nilai yang tinggi dan diterima. Nilai Coef\_G *absolute* phi ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan kreativiti dikekalkan untuk nilai coef\_G *absolute* phi ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Sosioemosi 1.* Dalam kajian-G, Coef\_G *relative* ( $\hat{\epsilon}_p^2$ ) memaparkan tahap perbezaan ralat *relative* varian (*relative error variance*). Dalam kajian-D Coef\_G *absolute* phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.78 menunjukkan perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.78

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Sosioemosi IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G <i>relative</i> ( $\hat{\epsilon}_p^2$ )	0.96	0.97	0.97	0.97	0.96	0.96
Coef_G <i>absolute</i> ( $\Phi$ )	0.90	0.91	0.91	0.90	0.90	0.90

Dalam kajian ini, *Coef\_G absolute phi* ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan sosioemosi dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.78, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, *Coef\_G absolute phi* ( $\Phi$ ) kekal pada 0.90, iaitu satu nilai yang tinggi dan diterima. Nilai *Coef\_G absolute phi* ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan sosioemosi dikekalkan untuk nilai *coef\_G absolute phi* ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Sosioemosi 2.* Dalam kajian-G, *Coef\_G relative* ( $\hat{E}_p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian-D *Coef\_G absolute phi* ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.79 menunjukkan perbezaan nilai kebolehpercayaan atau *Coef\_G* apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.79

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Sosioemosi IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.97	0.98	0.98	0.98	0.97	0.97
Coef_G absolute ( $\Phi$ )	0.97	0.97	0.97	0.97	0.97	0.97

Dalam kajian ini, Coef\_G absolute phi ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan sosioemosi dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.79, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambil kira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G absolute phi ( $\Phi$ ) kekal pada 0.97, iaitu satu nilai yang tinggi dan diterima. Nilai Coef\_G absolute phi ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan sosioemosi dikekalkan untuk nilai coef\_G absolute phi ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Rohani 1.* Dalam kajian-G, Coef\_G relative ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat relative varian (*relative error variance*). Dalam kajian-D Coef\_G absolute phi ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute*

*error variance*). Jadual 5.80 menunjukkan perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.80

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Rohani IPHKT 1*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.93	0.94	0.94	0.93	0.93	0.93
Coef_G <i>absolute</i> ( $\Phi$ )	0.78	0.78	0.78	0.78	0.78	0.78

Dalam kajian ini, Coef\_G *absolute* phi ( $\Phi$ ) sahaja diambilkira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan rohani dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.80, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G *absolute* phi ( $\Phi$ ) kekal pada 0.78, iaitu satu nilai yang masih boleh diterima. Nilai Coef\_G *absolute* phi ( $\Phi$ ) ini menghampiri nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas prestasi kepada domain perkembangan rohani dikekalkan untuk nilai coef\_G *absolute* phi ( $\Phi$ ) ini atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

*Domain Rohani 2*. Dalam kajian–G, Coef\_G relative ( $\hat{E}p^2$ ) memaparkan tahap perbezaan ralat *relative varian* (*relative error variance*). Dalam kajian–D Coef\_G *absolute phi* ( $\Phi$ ) menunjukkan darjah perbezaan secara mutlak varian ralat (*absolute error variance*). Jadual 5.81 menunjukkan perbezaan nilai kebolehpercayaan atau Coef\_G apabila bilangan kanak-kanak dan penilai ditambah atau diubahsuai.

Jadual 5.81

*Komponen Varians Kajian-D Berdasarkan Ubahsuai Bilangan Kanak-kanak dan Penilai Domain Rohani IPHKT 2*

	G-study	Opt.1	Opt. 2	Opt. 3	Opt. 4	Opt.5
Bilangan Kanak-kanak(K)	80	100	120	140	160	180
Bilangan Penilai(P:KI)	2	4	4	3	2	2
Coef_G relative ( $\hat{E}p^2$ )	0.95	0.96	0.96	0.96	0.95	0.95
Coef_G <i>absolute phi</i> ( $\Phi$ )	0.94	0.94	0.94	0.94	0.94	0.94

Dalam kajian ini, Coef\_G *absolute phi* ( $\Phi$ ) sahaja diambil kira kerana kajian ini adalah untuk mengkaji ralat varians terhadap skor kanak-kanak yang dinilai oleh dua penilai yang berbeza bagi domain perkembangan rohani dalam penilaian berasas prestasi di taska. Kajian ini juga membandingkan skor yang diberi oleh dua pemeriksa yang berbeza taska masing-masing. Daripada Jadual 5.81, didapati bilangan kanak-kanak yang sesuai dinilai dalam penilaian ini adalah seramai 80 orang, dengan mengambilkira bilangan pemeriksa dikekalkan kepada 2 orang. Dengan cadangan 80 orang ini, Coef\_G *absolute phi* ( $\Phi$ ) kekal pada 0.94, iaitu satu nilai yang tinggi dan diterima. Nilai Coef\_G *absolute phi* ( $\Phi$ ) ini melebihi nilai konvensional yang diterima iaitu 0.8. Keputusan untuk memilih bilangan kanak-kanak yang paling sesuai untuk dinilai dibuat berdasarkan pertimbangan terhadap faktor masa, kos, logistik dan lain-lain. Ini bermakna sekiranya bilangan kanak-kanak yang dinilai dikekalkan 80 orang, ia diterima dan mencukupi untuk menghadapi kekangan terhadap faktor masa, kos, logistik dan lain-lain. Oleh yang demikian, untuk kajian ini, penyelidik mencadangkan bilangan kanak-kanak seramai 80 dan penilai seramai 2 orang dalam penilaian berasas

prestasi kepada domain perkembangan rohani dikekalkan untuk nilai *coef\_G absolute phi* ( $\Phi$ ) atau nilai kebolehpercayaan yang tinggi selari dengan hasil dapatan ini.

**Ketaksamaan Kefungsian Faset (Differential Faset Functioning) Penilai dan Kumpulan Umur Kanak-kanak.** Model PRPF digunakan untuk menyiasat interaksi bias antara penilai dengan kumpulan umur. Istilah bias di antara penilai dan faset yang lain di sini bermaksud ketegasan atau kelonggaran penilai dalam membuat penilaian terhadap kanak-kanak mengikut kumpulan umur. Bias didefinisikan sebagai kecenderungan penilai dalam membuat penilaian yang rendah atau tinggi secara konsisten dari yang sepatutnya terhadap prestasi pelajar (Engelhard, 1994). Wingglesworth (1993) menyatakan analisis bias menentukan sistematik subpola tingkahlaku yang wujud dari interaksi antara penilai tertentu dengan aspek tertentu bagi situasi penilaian tertentu. Analisis bias dapat membantu penilai memahami proses penilaian yang berlaku dan dapat membuat penambahbaikan dalam pembentukan instrumen dan skala pengukuran.

Bagi kajian ini, analisis bias yang dilakukan adalah interaksi bias antara penilai dan kumpulan umur kanak-kanak. Persoalan kajian bagi interaksi bias dijawab mengikut setiap domain yang terdapat IPHKT 1 dan 2.

**Domain kognitif.** Jadual 5.82 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (severe) hingga mudah (leniency) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai

skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.82 menunjukkan 6 penilai (26, 32, 9, 25, 8, 24) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun. Manakala 10 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 26 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 2 tahun. Penilai ini menilai kumpulan umur kanak-kanak 2 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 338, lebih rendah daripada nilai jangkaan iaitu 357.63. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu  $-0.24$  ( $338-357.63$ ). Nilai ini menunjukkan  $0.24$  lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai  $0.24$  (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah  $-0.40$  logit dengan ketepatan (*standard error*, SE) iaitu  $0.13$ . Oleh yang demikian, penilai 26 adalah bersifat tegas menilai kumpulan umur kanak-kanak 2 tahun (bias= $-0.40$ ,  $t=-2.95$ ).

Seterusnya, penilai 26 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 26 adalah 443, manakala skor jangkaan adalah 423.37. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu  $0.32$ . Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah  $0.32$  logit dengan ketepatan (*standard error*, SE) iaitu  $0.13$ . Oleh yang demikian, penilai 26 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias=  $0.32$ ,  $t= 2.42$ ).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun (ukuran bias di antara 0.01 logit hingga 0.32 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah lebih bersifat tegas semasa memberi skor (ukuran bias di antara -0.02 logit hingga -0.40 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat mudah.

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan kognitif IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.42$ ) iaitu menunjukkan bahawa penilai 26 telah menilai kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 2 tahun dengan lebih tegas.



Jadual 5.82

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kognitif IPHKT 1

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Nilai Prob.
Penilai26	2	338	357.63	-0.24	-0.40	0.13	-2.95	0.0042
Penilai26	3	443	423.37	0.19	0.32	0.13	2.42	0.0173
Penilai 19	2	467	454.13	0.12	0.21	0.13	1.62	0.1074
Penilai 19	3	704	716.84	-0.08	-0.20	0.12	-1.66	0.0980
Penilai 32	2	541	556.64	-0.10	-0.13	0.09	-1.42	0.1587
Penilai 32	3	851	835.38	0.08	0.14	0.10	1.44	0.1521
Penilai 9	2	607	617.71	-0.07	-0.13	0.11	-1.19	0.2360
Penilai 9	3	828	817.29	0.06	0.12	0.11	1.12	0.2632
Penilai 31	2	551	538.01	0.08	0.10	0.09	1.16	0.2486
Penilai 31	3	804	817.01	-0.06	-0.10	0.09	-1.16	0.2482
Penilai 25	2	631	639.30	-0.06	-0.12	0.12	-1.03	0.3063
Penilai 25	3	862	853.70	0.04	0.12	0.12	0.99	0.3257
Penilai 8	2	594	602.31	-0.06	-0.09	0.10	-0.87	0.3854
Penilai 8	3	805	796.69	0.04	0.09	0.10	0.84	0.4012
Penilai 3	2	1008	998.26	0.05	0.15	0.13	1.17	0.2415
Penilai 3	3	1132	1141.75	-0.04	-0.07	0.09	-0.85	0.3941
Penilai 29	2	1212	1203.52	0.03	0.12	0.12	0.99	0.3244
Penilai 29	3	1267	1275.50	-0.03	-0.07	0.09	-0.80	0.4239
Penilai 4	2	1006	997.91	0.04	0.12	0.13	0.97	0.3314
Penilai 4	3	1133	1141.10	-0.03	-0.06	0.09	-0.71	0.4797
Penilai 13	2	657	652.89	0.03	0.08	0.14	0.55	0.5843
Penilai 13	3	830	834.11	-0.02	-0.05	0.11	-1.46	0.6468
Penilai 27	2	358	356.03	0.02	0.04	0.15	0.30	0.7687
Penilai 27	3	419	420.97	-0.02	-0.03	0.12	-1.24	0.8090
Penilai 23	2	492	490.04	0.02	0.03	0.12	0.23	0.8209
Penilai 23	3	638	639.93	-0.01	-0.03	0.12	-0.23	0.8176
Penilai 30	2	1204	1201.93	0.01	0.03	0.12	0.24	0.8115
Penilai 30	3	1271	1273.08	-0.01	-0.02	0.09	-0.20	0.8455
Penilai 22	2	672	670.92	0.01	0.03	0.16	0.17	0.8681
Penilai 22	3	864	865.08	-0.01	-0.02	0.12	-0.13	0.8936
Penilai 24	2	498	499.25	-0.01	-0.03	0.15	-0.18	0.8539
Penilai 24	3	1001	999.76	0.01	0.01	0.11	0.13	0.8931
Min		771.50	771.50	0	0.1	0.12	0.01	
s.p		273.5	273.36	0.07	0.01	0.02	1.10	

Jadual 5.83 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun IPHKT 2. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (severe) hingga mudah (leniency) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp*)

*Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.83 menunjukkan 7 penilai (31,29,26,4,30,9,27) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Manakala 8 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 31 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 4 tahun. Penilai ini menilai kumpulan umur kanak-kanak 4 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 1031, lebih rendah daripada nilai jangkaan iaitu 1066.20. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.14 (1031-1066.20). Nilai ini menunjukkan 0.14 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.14 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.37 logit dengan ketepatan (*standard error*, SE) iaitu 0.13. Oleh yang demikian, penilai 31 adalah bersifat tegas menilai kumpulan umur kanak-kanak 4 tahun (bias=-0.37, t=-3.71).

Seterusnya, penilai 31 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 31 adalah 1108, manakala skor jangkaan adalah 1072.95. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.14. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.42 logit dengan ketepatan (*standard error*, SE) iaitu

0.11. Oleh yang demikian, penilai 31 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.42, t= 3.72).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun (ukuran bias di antara 0.01 logit hingga 0.42 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah lebih bersifat tegas semasa memberi skor (ukuran bias di antara -0.02 logit hingga -0.37 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat mudah.

Jadual 5.83

*Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kognitif IPHKT 2*

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 31	3	1108	1072.95	0.14	0.42	0.11	3.72	0.0002
Penilai 31	4	1031	1066.20	-0.14	-0.37	0.10	-3.71	0.0003
Penilai 29	3	1766	1745.33	0.05	0.27	0.12	2.31	0.0213
Penilai 29	4	673	693.70	-0.13	-0.41	0.14	-3.03	0.0028
Penilai 32	3	1146	1164.80	-0.08	-0.32	0.13	-2.54	0.0117
Penilai 32	4	1179	1160.26	0.07	0.36	0.15	2.49	0.0134
Penilai 19	3	862	875.74	-0.07	-0.36	0.15	-2.31	0.0219
Penilai 19	4	909	895.26	0.07	0.66	0.25	2.69	0.0079
Penilai 11	3	1169	1180.65	-0.05	-0.23	0.14	-1.68	0.0941
Penilai 11	4	1816	1804.36	0.03	0.20	0.14	1.49	0.1368
Penilai 15	3	1165	1175.37	-0.04	-0.19	0.13	-1.45	0.1471
Penilai 15	4	1808	1797.64	0.03	0.17	0.13	1.29	0.1990
Penilai 24	3	1283	1294.82	-0.04	-0.16	0.11	-1.39	0.1642
Penilai 24	4	900	888.22	0.06	0.48	0.22	2.19	0.0300
Penilai 26	3	470	461.67	0.07	0.14	0.13	1.07	0.2875
Penilai 26	4	520	528.37	-0.07	-0.19	0.15	-1.28	0.2046
Penilai 4	3	1617	1610.06	0.02	0.09	0.12	0.79	0.4291
Penilai 4	4	1779	1785.55	-0.02	-0.11	0.13	-0.86	0.3883
Penilai 1	3	1256	1263.29	-0.03	-0.08	0.11	-0.79	0.4304
Penilai 1	4	886	878.75	0.04	0.21	0.17	1.19	0.2371
Penilai 30	3	1809	1805.50	0.01	0.06	0.14	0.47	0.6388
Penilai 30	4	722	725.51	-0.02	-0.11	0.17	-0.63	0.5307
Penilai 20	3	882	885.06	-0.02	-0.10	0.18	-0.56	0.5781
Penilai 20	4	905	901.93	0.02	0.15	0.23	0.66	0.5072
Penilai 9	3	1096	1093.19	0.01	0.04	0.11	0.31	0.7541
Penilai 9	4	892	894.76	-0.01	-0.10	0.19	-0.54	0.5927
Penilai 27	3	489	486.39	0.02	0.05	0.14	0.35	0.7264
Penilai 27	4	543	545.64	-0.02	-0.07	0.16	-0.43	0.6697
Penilai 3	3	1571	1574.57	-0.01	-0.04	0.10	-0.37	0.7116
Penilai 3	4	1763	1758.96	0.01	0.06	0.12	0.48	0.6294
Penilai 8	3	1076	1075.26	0.00	0.01	0.11	0.08	0.9370
Penilai 8	4	888	888.68	0.00	-0.02	0.18	-0.12	0.9021
Min		1124.3	1124.33	0	0.2	0.15	0.00	
s.p		414.8	412.18	0.06	0.25	0.04	1.69	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan kognitif IPHKT 2, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=3.72$ ) iaitu menunjukkan bahawa penilai 31 telah menilai kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 4 tahun dengan lebih tegas.

**Domain bahasa.** Jadual 5.84 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.84 menunjukkan 7 penilai (25,31,9,3,24,29,8) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Manakala 9 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun. Penilai 25 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 3 tahun. Penilai ini menilai kumpulan

umur kanak-kanak 3 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 614, lebih rendah daripada nilai jangkaan iaitu 629.89. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu  $-0.12$  ( $614-629.89$ ). Nilai ini menunjukkan  $0.12$  lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai  $0.12$  (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah  $-0.30$  logit dengan ketepatan (*standard error*, SE) iaitu  $0.13$ . Oleh yang demikian, penilai 25 adalah bersifat tegas menilai kumpulan umur kanak-kanak 3 tahun (bias= $-0.30$ ,  $t=-2.29$ )

Seterusnya, penilai 25 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 2 tahun iaitu skor sebenar penilai 25 adalah 589, manakala skor jangkaan adalah 573.10. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu  $0.12$ . Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah  $0.17$  logit dengan ketepatan (*standard error*, SE) iaitu  $0.11$ . Oleh yang demikian, penilai 25 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 2 tahun (bias=  $0.17$ ,  $t=1.57$ ).

Rumusannya, penilai-penilai adalah lebih bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun (ukuran bias di antara  $0.01$  logit hingga  $0.17$  logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 2 tahun, penilai-penilai adalah lebih bersifat tegas semasa memberi skor (ukuran bias di antara  $-0.02$  logit hingga  $-0.30$  logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat mudah.

Jadual 5.84

*Laporan bias guru penilai dan kumpulan umur kanak-kanak domain bahasa IPHKT 1*

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob.
Penilai 25	2	589	573.10	0.12	0.17	0.11	1.57	0.1191
Penilai 25	3	614	629.89	-0.12	-0.30	0.13	-2.29	0.0233
Penilai 31	2	497	477.05	0.14	0.12	0.08	1.55	0.1242
Penilai 31	3	520	539.93	-0.14	-0.13	0.08	-1.65	0.1014
Penilai 32	2	471	448.28	-0.12	-0.10	0.08	-1.35	0.1787
Penilai 32	3	567	549.71	0.12	0.13	0.09	1.46	0.1475
Penilai 9	2	568	557.81	0.08	0.10	0.10	0.98	0.3285
Penilai 9	3	568	578.18	-0.08	-0.13	0.11	-1.16	0.2482
Penilai 23	2	477	486.51	-0.08	-0.10	0.10	-1.00	0.3173
Penilai 23	3	429	419.48	0.10	0.15	0.13	1.15	0.2526
Penilai 4	2	844	854.78	-0.05	-0.09	0.09	-0.98	0.3298
Penilai 4	3	866	855.21	0.06	0.15	0.12	1.21	0.2278
Penilai 22	2	614	618.47	-0.03	-0.08	0.13	-0.60	0.5476
Penilai 22	3	610	605.53	0.03	0.08	0.13	0.57	0.5666
Penilai 3	2	837	831.28	0.03	0.04	0.09	0.48	0.6288
Penilai 3	3	874	879.71	-0.03	-0.06	0.10	-0.57	0.5682
Penilai 26	2	303	307.79	-0.06	-0.05	0.10	-0.50	0.6168
Penilai 26	3	247	242.20	0.08	0.07	0.12	0.58	0.5633
Penilai 24	2	509	505.60	0.03	0.03	0.09	0.32	0.7531
Penilai 24	3	652	655.39	-0.02	-0.05	0.12	-0.42	0.6750
Penilai 30	2	914	919.74	-0.02	-0.03	0.07	-0.39	0.6983
Penilai 30	3	969	963.25	0.03	0.05	0.10	0.55	0.5853
Penilai 27	2	303	306.55	-0.04	-0.04	0.10	-0.37	0.7110
Penilai 27	3	236	232.43	0.06	0.05	0.12	0.44	0.6650
Penilai 19	2	425	428.37	-0.03	-0.03	0.10	-0.34	0.7310
Penilai 19	3	452	448.62	0.03	0.04	0.11	0.38	0.7077
Penilai 29	2	927	923.74	0.01	0.02	0.07	0.22	0.8247
Penilai 29	3	962	965.26	-0.02	-0.03	0.09	-0.31	0.7549
Penilai 8	2	549	547.30	0.01	0.01	0.09	0.16	0.8747
Penilai 8	3	569	570.68	-0.01	-0.02	0.11	-0.19	0.8535
Penilai 13	2	571	571.94	-0.01	-0.02	0.13	-0.12	0.9008
Penilai 13	3	645	644.06	0.01	0.01	0.12	0.12	0.9084
Min		599.3	599.31	0	0	0.10	-0.02	
s.p		203.1	203.33	0.07	0.10	0.02	0.92	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan bahasa IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=-2.29$ ) iaitu menunjukkan bahawa penilai 25 telah menilai

kumpulan umur kanak-kanak 3 tahun dengan lebih tegas. dan menilai kumpulan umur kanak-kanak 2 tahun dengan lebih pemurah (*lenient*).

Jadual 5.85 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.85 menunjukkan 8 penilai (31,29,26,15,4,27,8,9) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Manakala 8 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 31 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 4 tahun. Penilai ini menilai kumpulan umur kanak-kanak 4 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 844, lebih rendah daripada nilai jangkaan iaitu 885.48. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.19 (844-885.48). Nilai ini menunjukkan 0.19 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.19 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.24 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.07. Oleh yang demikian, penilai adalah bersifat tegas menilai kumpulan umur kanak-kanak 4 tahun (bias=-0.24, t=-3.22).

Seterusnya, penilai 31 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 31 adalah 904, manakala skor jangkaan adalah 862.50. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.19. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.27 logit dengan ketepatan (*standard error*, SE) iaitu 0.08. Oleh yang demikian, penilai 31 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias=0.27, t=3.25).

Rumusannya, penilai-penilai adalah mempunyai sifat ketegasan yang sama semasa memberi skor kepada kumpulan umur kanak-kanak 3 dan 4 tahun iaitu seramai jumlah penilai adalah seimbang iaitu 8 penilai bersifat tegas dan 8 penilai bersifat mudah bagi kedua-dua kumpulan umur.



Jadual 5.85

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain bahasa IPHKT 2

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 31	3	904	862.50	0.19	0.27	0.08	3.25	0.0014
Penilai 31	4	844	885.48	-0.19	-0.24	0.07	-3.22	0.0015
Penilai 32	3	941	967.59	-0.12	-0.24	0.09	-2.62	0.0094
Penilai 32	4	1014	987.41	0.12	0.33	0.12	2.74	0.0066
Penilai 29	3	1079	1059.33	0.08	0.16	0.09	1.72	0.0861
Penilai 29	4	572	591.66	-0.14	-0.22	0.10	-2.14	0.0340
Penilai 11	3	896	916.33	-0.09	-0.14	0.08	-1.73	0.0848
Penilai 11	4	1538	1517.67	0.06	0.23	0.11	2.07	0.0395
Penilai 26	3	382	370.06	0.11	0.10	0.09	1.10	0.2759
Penilai 26	4	446	457.91	-0.11	-0.14	0.11	-1.34	0.1845
Penilai 3	3	1265	1282.12	-0.06	-0.09	0.07	-1.28	0.1999
Penilai 3	4	1494	1477.30	0.05	0.15	0.10	1.53	0.1263
Penilai 1	3	1006	1021.24	-0.06	-0.09	0.08	-1.18	0.2388
Penilai 1	4	765	749.74	0.09	0.26	0.14	1.88	0.0623
Penilai 15	3	940	930.69	0.04	0.07	0.09	0.82	0.4149
Penilai 15	4	1523	1532.31	-0.03	-0.10	0.10	-0.99	0.3226
Penilai 24	3	1043	1054.35	-0.05	-0.08	0.08	-0.95	0.3453
Penilai 24	4	774	762.64	0.07	0.23	0.15	1.53	0.1288
Penilai 4	3	1310	1304.38	0.02	0.04	0.08	0.44	0.6590
Penilai 4	4	1482	1488.01	-0.02	-0.05	0.09	-0.57	0.5672
Penilai 27	3	406	401.76	0.04	0.04	0.10	0.41	0.6804
Penilai 27	4	474	478.23	-0.04	-0.06	0.12	-0.53	0.6004
Penilai 8	3	903	899.08	0.02	0.02	0.07	0.27	0.7880
Penilai 8	4	743	746.77	-0.02	-0.05	0.12	-0.46	0.6492
Penilai 30	3	1116	1119.39	-0.01	-0.04	0.10	-0.35	0.7234
Penilai 30	4	632	628.61	0.02	0.06	0.14	0.46	0.6451
Penilai 19	3	722	724.06	-0.01	-0.02	0.11	-0.22	0.8223
Penilai 19	4	768	765.94	0.01	0.04	0.14	0.28	0.7775
Penilai 20	3	674	675.72	-0.01	-0.01	0.09	-0.16	0.8727
Penilai 20	4	736	734.27	0.01	0.02	0.11	0.19	0.8457
Penilai 9	3	849	847.67	0.01	0.01	0.07	0.09	0.9244
Penilai 9	4	747	748.19	-0.01	-0.02	0.12	-0.15	0.8846
Min		905.9	905.89	0	0.01	0.10	0.03	
s.p		316.6	315.46	0.08	0.15	0.02	1.49	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan kognitif IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=3.25$ ) iaitu menunjukkan bahawa penilai 31 telah

menilai kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 4 tahun dengan lebih tegas.

**Domain fizikal.** Jadual 5.86 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.86 menunjukkan 8 penilai (22,24,13,23,4,27,9,32) daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun. Manakala 7 penilai daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 22 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 2 tahun. Penilai ini menilai kumpulan umur kanak-kanak 2 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 890, lebih rendah daripada nilai jangkaan iaitu 917.42. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.14 (890-917.42). Nilai ini menunjukkan 0.14 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.14 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.51 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.12. Oleh yang demikian, penilai 22 adalah bersifat tegas menilai kumpulan umur kanak-kanak 2 tahun (bias=-0.51, t=-4.18).

Seterusnya, penilai 22 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 22 adalah 1369, manakala skor jangkaan adalah 1341.58. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.10. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.44 logit dengan ketepatan (*standard error*, SE) iaitu 0.14. Oleh yang demikian, penilai 22 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.44, t=3.16).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun (ukuran bias di antara 0.00 logit hingga 0.44 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 2 tahun, penilai-penilai adalah bersifat seimbang semasa memberi skor dan mempunyai ketegasan yang sama iaitu 8 penilai bersifat tegas dan 8 penilai bersifat mudah.

Jadual 5.86

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain fizikal IPHKT 1

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 22	2	890	917.42	-0.14	-0.51	0.12	-4.18	0.0000
Penilai 22	3	1369	1341.58	0.10	0.44	0.14	3.16	0.0018
Penilai 24	2	750	758.92	-0.06	-0.23	0.15	-1.50	0.1344
Penilai 24	3	1538	1529.09	0.03	0.16	0.14	1.66	0.2489
Penilai 13	2	921	927.29	-0.03	-0.16	0.15	-1.03	0.3057
Penilai 13	3	1362	1355.72	0.02	0.10	0.13	0.77	0.4443
Penilai 30	2	1591	1576.02	0.05	0.42	0.19	2.25	0.0251
Penilai 30	3	1941	1955.99	-0.04	-0.13	0.09	-1.40	0.1611
Penilai 29	2	1592	1577.48	0.04	0.42	0.19	2.22	0.0272
Penilai 29	3	1955	1969.53	-0.03	-0.12	0.09	-1.38	0.1693
Penilai 27	2	277	279.37	-0.04	-0.11	0.21	-0.53	0.6015
Penilai 27	3	655	652.64	0.02	0.03	0.11	0.26	0.7942
Penilai 23	2	664	672.84	-0.06	-0.11	0.11	-0.97	0.3315
Penilai 23	3	1056	1047.11	0.04	0.21	0.16	1.30	0.1945
Penilai 31	2	873	856.23	0.07	0.09	0.08	1.25	0.2120
Penilai 31	3	1215	1231.80	-0.06	-0.10	0.08	-1.34	0.1817
Penilai 4	2	1415	1419.34	-0.01	-0.10	0.15	-0.68	0.4951
Penilai 4	3	1771	1766.67	0.01	0.03	0.08	0.36	0.7173
Penilai 8	2	823	817.26	0.03	0.05	0.10	0.55	0.5807
Penilai 8	3	1315	1320.76	-0.02	-0.06	0.10	-0.58	0.5611
Penilai 9	2	823	828.06	-0.03	-0.05	0.10	-0.50	0.6178
Penilai 9	3	1331	1325.96	0.02	0.06	0.11	0.53	0.5954
Penilai 3	2	1406	1997.02	0.03	0.17	0.14	1.19	0.2345
Penilai 3	3	1687	1695.99	-0.02	-0.05	0.07	-0.65	0.5136
Penilai 25	2	882	879.98	0.01	0.03	0.12	0.24	0.8120
Penilai 25	3	1355	1357.03	-0.01	-0.03	0.12	-0.25	0.8014
Penilai 26	2	281	279.66	0.02	0.07	0.23	0.30	0.7652
Penilai 26	3	649	650.35	-0.01	-0.02	0.11	-0.15	0.8802
Penilai 32	2	898	899.20	-0.01	-0.01	0.08	-0.09	0.9253
Penilai 32	3	1260	1258.82	0.00	0.01	0.09	0.10	0.9198
Penilai 19	2	675	675.40	0.00	0.00	0.11	-0.04	0.9642
Penilai 19	3	1049	1048.54	0.00	0.01	0.15	0.07	0.9460
Min		1133.4	1133.41	0	0.02	0.12	0.01	
s.p		434.0	433.46	0.04	0.18	0.04	1.33	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan fizikal IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.42$ ) iaitu menunjukkan bahawa penilai 26 telah menilai

kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 2 tahun dengan lebih tegas.

Jadual 5.87 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.87 menunjukkan 6 penilai (4,29,31,26,9,27,) daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Manakala 9 penilai daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 4 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 4 tahun. Penilai ini menilai kumpulan umur kanak-kanak 4 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 1347, lebih rendah daripada nilai jangkaan iaitu 1372.56. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.09 (1347-1372.56). Nilai ini menunjukkan 0.09 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.09 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.37 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.11. Oleh yang demikian, penilai 4 adalah bersifat tegas menilai kumpulan umur kanak-kanak 4 tahun (bias=-0.37,  $t=-3.26$ ).

Seterusnya, penilai 4 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 4 adalah 1254, manakala skor jangkaan adalah 1228.42. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.10. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.34 logit dengan ketepatan (*standard error*, SE) iaitu 0.12. Oleh yang demikian, penilai 4 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.34,  $t= 2.78$ ).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun (ukuran bias di antara 0.01 logit hingga 0.41 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah lebih bersifat tegas semasa memberi skor (ukuran bias di antara -0.03 logit hingga -0.37 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat mudah.

Jadual 5.87

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain fizikal IPHKT 2

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 4	3	1254	1228.42	0.10	0.34	0.12	2.78	0.0058
Penilai 4	4	1347	1372.56	-0.09	-0.37	0.11	-3.26	0.0013
Penilai29	3	1005	989.06	0.07	0.28	0.14	2.07	0.0452
Penilai29	4	523	538.93	-0.14	-0.42	0.15	-2.73	0.0074
Penilai3	3	1176	1201.98	-0.10	-0.25	0.10	-2.64	0.0087
Penilai 3	4	1382	1355.99	0.09	0.41	0.13	3.07	0.0023
Penilai31	3	863	845.12	0.09	0.26	0.12	2.08	0.0386
Penilai31	4	861	878.86	-0.09	-0.28	0.12	-2.31	0.0222
Penilai19	3	536	545.39	-0.08	-0.32	0.18	-1.82	0.0709
Penilai19	4	818	808.61	0.06	0.49	0.25	1.93	0.0556
Penilai26	3	372	361.73	0.11	0.19	0.14	1.38	0.1717
Penilai26	4	402	412.26	-0.11	-0.30	0.16	-1.81	0.0739
Penilai20	3	537	546.13	-0.08	-0.32	0.18	-1.79	0.0762
Penilai20	4	814	804.87	0.06	0.50	0.26	1.92	0.0572
Penilai15	3	463	872.76	-0.05	-0.15	0.12	-1.25	0.2123
Penilai15	4	1401	1391.23	0.03	0.20	0.15	1.35	0.1797
Penilai24	3	966	975.66	-0.04	-0.12	0.11	-1.11	0.2676
Penilai24	4	695	685.33	0.07	0.76	0.33	2.30	0.0230
Penilai9	3	866	863.12	0.02	0.04	0.12	0.35	0.7240
Penilai9	4	683	685.87	-0.02	-0.11	0.20	-0.58	0.5616
Penilai27	3	368	364.84	0.03	0.06	0.13	0.42	0.6754
Penilai27	4	409	412.14	-0.03	-0.10	0.17	-0.55	0.5807
Penilai11	3	877	880.80	-0.02	-0.06	0.13	-0.49	0.6217
Penilai11	4	1403	1399.19	0.01	0.08	0.14	0.54	0.5924
Penilai1	3	955	959.21	-0.02	-0.05	0.11	-0.46	0.6473
Penilai1	4	695	690.78	0.03	0.22	0.24	0.93	0.3529
Penilai8	3	845	847.61	-0.01	-0.03	0.11	-0.30	0.7625
Penilai8	4	674	671.37	0.02	0.10	0.20	0.50	0.6153
Penilai32	3	875	876.98	-0.01	-0.03	0.13	-0.26	0.7929
Penilai32	4	911	909.01	0.01	0.05	0.15	0.30	0.7647
Penilai30	3	1017	1016.49	0.00	0.01	0.15	0.08	0.9369
Penilai30	4	558	558.53	0.00	-0.02	0.20	-0.11	0.9154
Min		842.2	842.21	0	0.03	0.16	0.01	
s.p		298.1	296.29	0.06	0.28	0.05	1.66	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan fizikal IPHKT 2, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.78$ ) iaitu menunjukkan bahawa penilai 4 telah menilai

kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 4 tahun dengan lebih tegas.

**Domain kreativiti.** Jadual 5.88 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.88 menunjukkan 7 penilai (31,29,3,13,19,3,24) daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Manakala 8 penilai daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 31 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 3 tahun. Penilai ini menilai kumpulan umur kanak-kanak 3 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 617, lebih rendah daripada nilai jangkaan iaitu 647.86. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.20 (617-647.86). Nilai ini menunjukkan 0.20 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.20 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.29 logit dengan ketepatan (*standard*



*error*, SE) iaitu 0.09. Oleh yang demikian, penilai 31 adalah bersifat tegas menilai kumpulan umur kanak-kanak 3 tahun (bias=-0.29, t=-3.14)

Seterusnya, penilai 31 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 2 tahun iaitu skor sebenar penilai 31 adalah 550, manakala skor jangkaan adalah 519.17. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.23. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.27 logit dengan ketepatan (*standard error*, SE) iaitu 0.10. Oleh yang demikian, penilai 31 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.27, t= 2.80).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun (ukuran bias di antara 0.10 logit hingga 0.57 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 2 tahun, penilai-penilai mempunyai ketegasan yang sama iaitu seimbang jumlah penilai yang tegas dan yang bersifat mudah semasa memberi skor.

Jadual 5.88

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kreativiti IPHKT 1

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 31	2	550	519.17	0.23	0.27	0.10	2.80	0.0059
Penilai 31	3	617	647.86	-0.20	-0.29	0.09	-3.14	0.0020
Penilai 29	2	277	245.91	0.56	0.47	0.57	4.30	0.0001
Penilai 29	3	819	850.10	-0.16	-0.29	0.09	-3.13	0.0020
Penilai 32	2	469	499.82	-0.23	-0.23	0.09	-2.71	0.0075
Penilai 32	3	663	632.21	0.20	0.32	0.11	2.98	0.0032
Penilai 22	2	570	581.24	-0.09	-0.24	0.14	-1.73	0.0863
Penilai 22	3	752	740.76	0.07	0.18	0.13	1.39	0.1666
Penilai 30	2	228	237.60	-0.17	-0.24	0.15	-1.59	0.1186
Penilai 30	3	833	823.42	0.05	0.08	0.10	0.89	0.3770
Penilai 8	2	532	541.83	-0.08	-0.13	0.11	-1.15	0.2525
Penilai 8	3	717	707.17	0.06	0.11	0.11	1.02	0.3097
Penilai4	2	429	434.97	-0.06	-0.13	0.14	-0.91	0.3637
Penilai4	3	1060	1054.04	0.03	0.06	0.10	0.60	0.5461
Penilai9	2	544	551.05	-0.06	-0.10	0.12	-0.87	0.3883
Penilai 9	3	726	718.94	0.04	0.09	0.11	0.77	0.4448
Penilai 25	2	474	466.49	0.07	0.15	0.15	1.03	0.3040
Penilai25	3	820	827.51	-0.04	-0.10	0.11	-0.86	0.3904
Penilai26	2	178	181.49	-0.08	-0.10	0.17	-0.59	0.5555
Penilai26	3	348	344.51	0.04	0.06	0.13	0.46	0.6474
Penilai 13	2	584	579.27	0.04	0.11	0.16	0.71	0.4763
Penilai13	3	733	437.72	-0.03	-0.07	0.12	-0.57	0.5662
Penilai19	2	365	362.77	0.03	0.04	0.13	0.30	0.7664
Penilai19	3	671	673.19	-0.02	-0.07	0.17	-0.39	0.6977
Penilai23	2	413	415.37	-0.02	-0.03	0.11	-0.27	0.7868
Penilai23	3	571	568.57	0.02	0.06	0.15	0.36	0.7172
Penilai3	2	435	433.13	0.02	0.04	0.15	0.28	0.7797
Penilai3	3	1048	1049.87	-0.01	-0.02	0.10	-0.19	0.8518
Penilai24	2	433	431.75	0.01	0.03	0.16	0.20	0.8450
Penilai24	3	477	478.25	-0.01	-0.02	0.11	-0.14	0.8859
Penilai27	2	186	185.60	0.01	0.01	0.18	0.07	0.9441
Penilai27	3	351	351.40	0.00	-0.01	0.14	-0.05	0.9567
Min		571.0	571.03	0.01	0.06	0.14	0	
s.p		225.7	226.58	0.14	0.46	0.08	1.57	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan kreativiti IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.80$ ) iaitu menunjukkan bahawa penilai 31 telah

menilai kumpulan umur kanak-kanak 2 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 3 tahun dengan lebih tegas.

Jadual 5.89 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata  $S-J$  (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.89 menunjukkan 6 penilai (31,29,26,3,27,19) daripada 14 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Manakala 8 penilai daripada 14 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 31 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 4 tahun. Penilai ini menilai kumpulan umur kanak-kanak 4 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 952, lebih rendah daripada nilai jangkaan iaitu 982.05. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu  $-0.13(952-982.05)$ . Nilai ini menunjukkan 0.13 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.13 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah  $-0.25$  logit dengan ketepatan (*standard error, SE*)

iaitu 0.09. Oleh yang demikian, penilai 31 adalah bersifat tegas menilai kumpulan umur kanak-kanak 4 tahun (bias=-0.25, t=-2.84).

Seterusnya, penilai 31 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 31 adalah 945, manakala skor jangkaan adalah 914.93. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.13. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.23 logit dengan ketepatan (*standard error*, SE) iaitu 0.09. Oleh yang demikian, penilai 31 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.23, t=2.57).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun (ukuran bias di antara 0.00 logit hingga 0.23 logit) iaitu seramai 10 penilai berbanding 6 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah bersifat lebih seimbang semasa memberi skor kerana 8 penilai yang bersifat mudah dan 8 penilai yang lebih bertegas kepada kedua-dua peringkat umur kanak-kanak tersebut.

Jadual 5.89

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain kreativiti IPHKT 2

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 31	3	945	914.93	0.13	0.23	0.09	2.57	0.0109
Penilai 31	4	952	982.05	-0.13	-0.25	0.09	-2.84	0.0049
Penilai29	3	1485	1465.05	0.06	0.16	0.09	1.76	0.0801
Penilai29	4	628	647.94	-0.13	-0.26	0.11	-2.36	0.0196
Penilai26	3	458	440.48	0.14	0.17	0.10	1.73	0.0870
Penilai26	4	510	527.51	-0.14	-0.25	0.12	-2.17	0.0318
Penilai15	3	1052	1071.68	-0.08	-0.18	0.09	-1.92	0.0561
Penilai 15	4	1582	1562.30	0.06	0.31	0.13	2.32	0.0208
Penilai20	3	610	621.17	-0.08	-0.12	0.10	-1.19	0.2358
Penilai20	4	933	921.83	0.06	0.19	0.13	1.39	0.1663
Penilai24	3	1070	1081.77	-0.05	-0.09	0.08	-1.02	0.3103
Penilai24	4	832	820.22	0.07	0.31	0.17	1.78	0.0763
Penilai 1	3	1082	1093.52	-0.05	-0.09	0.09	-1.01	0.3132
Penilai 1	4	834	822.47	0.07	0.31	0.18	1.78	0.0766
Penilai3	3	1362	1353.13	0.03	0.07	0.09	0.78	0.4345
Penilai3	4	1534	1542.86	-0.03	-0.10	0.10	-0.96	0.3387
Penilai8	3	897	909.98	-0.05	-0.07	0.07	-0.93	0.3546
Penilai8	4	879	866.02	0.07	0.20	0.13	1.56	0.1214
Penilai11	3	954	962.95	-0.04	-0.07	0.09	-0.81	0.4169
Penilai11	4	1549	1540.03	0.03	0.11	0.11	0.95	0.3411
Penilai27	3	443	436.81	0.05	0.06	0.10	0.61	0.5446
Penilai27	4	519	525.19	-0.05	-0.09	0.12	-0.76	0.4481
Penilai32	4	1037	1029.30	0.03	0.10	0.11	0.85	0.3943
Penilai32	3	968	975.68	-0.03	-0.07	0.09	-0.73	0.4646
Penilai19	3	663	658.15	0.03	0.08	0.13	0.60	0.5509
Penilai19	4	943	947.85	-0.02	-0.10	0.14	-0.73	0.4686
Penilai9	3	878	881.18	-0.01	-0.02	0.08	-0.25	0.8064
Penilai9	4	807	803.82	0.02	0.06	0.14	0.43	0.6702
Penilai30	3	1508	1509.30	0.00	-0.01	0.10	-0.13	0.8994
Penilai30	4	673	671.69	0.01	0.02	0.14	0.18	0.8602
Penilai4	3	1371	1370.60	0.00	0.00	0.09	0.04	0.9704
Penilai4	4	1554	1554.39	0.00	0.00	0.11	-0.04	0.9650
Min		571.0	571.03	0.01	0.06	0.14	0	
s.p		225.7	226.58	0.14	0.46	0.08	1.57	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan kreativiti IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.42$ ) iaitu menunjukkan bahawa penilai 26 telah

menilai kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 2 tahun dengan lebih tegas.

**Domain sosioemosi.** Jadual 5.90 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.90 menunjukkan 7 penilai (23,22,32,8,24,4,26) daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun. Manakala 8 penilai daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 23 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 2 tahun. Penilai ini menilai kumpulan umur kanak-kanak 2 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 685, lebih rendah daripada nilai jangkaan iaitu 743.86. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.26 (685-743.86). Nilai ini menunjukkan 0.26 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.26 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.20 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.06. Oleh yang demikian, penilai 23 adalah bersifat tegas menilai kumpulan umur kanak-kanak 2 tahun (bias=-0.20, t=-3.51).

Seterusnya, penilai 23 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 23 adalah 854, manakala skor jangkaan adalah 795.13. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.29. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.44 logit dengan ketepatan (*standard error*, SE) iaitu 0.09. Oleh yang demikian, penilai 23 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.44, t=4.62).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun (ukuran bias di antara 0.00 logit hingga 0.44 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah bersifat lebih seimbang semasa memberi skor kerana 8 penilai yang bersifat mudah dan 8 penilai yang lebih bertegas kepada kedua-dua peringkat umur kanak-kanak tersebut.

Jadual 5.90

*Laporan bias guru penilai dan kumpulan umur kanak-kanak domain sosioemosi IPHKT 1*

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 23	2	685	743.86	-0.26	-0.20	0.06	-3.51	0.0005
Penilai 23	3	854	795.13	0.29	0.44	0.09	4.62	0.0000
Penilai 31	2	1020	974.16	0.15	0.13	0.05	2.38	0.0181
Penilai 31	3	1165	1210.84	-0.15	-0.18	0.06	-2.97	0.0032
Penilai 22	2	1155	1189.22	-0.13	-0.23	0.08	-2.93	0.0037
Penilai 22	3	1290	1255.77	0.12	0.26	0.09	2.86	0.0045
Penilai 25	2	1052	1020.94	0.12	0.15	0.07	2.13	0.0098
Penilai 25	3	1323	1354.05	-0.10	-0.20	0.08	-2.60	0.0344
Penilai 32	2	923	968.61	-0.15	-0.12	0.05	-2.35	0.0194
Penilai 32	3	1255	1209.39	0.15	0.20	0.07	2.93	0.0037
Penilai 3	2	1870	1834.54	0.08	0.14	0.06	2.16	0.0310
Penilai 3	3	1769	1804.45	-0.08	-0.11	0.06	-2.04	0.0416
Penilai 19	2	803	780.43	0.11	0.11	0.07	1.53	0.1286
Penilai 19	3	877	899.56	-0.11	-0.16	0.08	-0.97	0.0498
Penilai 8	2	951	967.05	-0.06	-0.05	0.06	-0.93	0.3556
Penilai 8	3	1216	1199.94	0.06	0.09	0.07	1.15	0.2508
Penilai 30	2	2285	2269.59	0.03	0.05	0.06	0.91	0.3622
Penilai 30	3	1976	1991.41	-0.03	-0.05	0.06	-0.88	0.3805
Penilai 27	2	558	549.81	0.06	0.05	0.08	0.66	0.5106
Penilai 27	3	419	427.18	-0.07	-0.07	0.09	-0.79	0.4307
Penilai 24	2	915	922.16	-0.03	-0.06	0.09	-0.66	0.5131
Penilai 24	3	1605	1597.84	0.02	0.06	0.09	0.64	0.5199
Penilai 9	2	969	960.82	0.03	0.03	0.06	0.47	0.6361
Penilai 9	3	1198	1206.16	-0.03	-0.04	0.07	-0.59	0.5586
Penilai 4	2	1795	1802.09	-0.02	-0.02	0.06	-0.42	0.6745
Penilai 4	3	1824	1816.90	0.02	0.02	0.05	0.39	0.6991
Penilai 29	2	2268	2264.47	0.01	0.01	0.06	0.21	0.8343
Penilai 29	3	1980	1983.54	-0.01	-0.01	0.06	-0.20	0.8401
Penilai 26	2	546	548.18	-0.02	-0.02	0.09	-0.20	0.8435
Penilai 26	3	462	519.81	0.02	0.03	0.11	0.24	0.8071
Penilai 13	2	1188	1187.45	0.00	0.00	0.09	0.05	0.9615
Penilai 13	3	1263	1263.54	0.00	0.00	0.09	-0.05	0.9632
Min		1233.1	1233.09	0	0.01	0.07	0.01	
s.p		505.5	504.34	0.11	0.14	0.02	1.87	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan sosioemosi IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=4.62$ ) iaitu menunjukkan bahawa penilai 23 telah



menilai kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 2 tahun dengan lebih tegas.

Jadual 5.91 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.91 menunjukkan 7 penilai (32,19,11,3,9,24,1) daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Manakala 8 penilai daripada 15 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Penilai 32 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 3 tahun. Penilai ini menilai kumpulan umur kanak-kanak 3 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 988, lebih rendah daripada nilai jangkaan iaitu 1029.88. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.18 (988-1029.88). Nilai ini menunjukkan 0.18 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.18 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.42 logit dengan ketepatan (*standard error*, SE)

iaitu 0.09. Oleh yang demikian, penilai 32 adalah bersifat tegas menilai kumpulan umur kanak-kanak 3 tahun (bias=-0.42, t=-4.47)

Seterusnya, penilai 32 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 4 tahun iaitu skor sebenar penilai 32 adalah 1090, manakala skor jangkaan adalah 1048.11. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.18. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.68 logit dengan ketepatan (*standard error, SE*) iaitu 0.14. Oleh yang demikian, penilai 32 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 4 tahun (bias= 0.68, t=4.79).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun (ukuran bias di antara 0.00 logit hingga 0.68 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 4 tahun, penilai-penilai adalah bersifat lebih seimbang semasa memberi skor kerana 8 penilai yang bersifat mudah dan 8 penilai yang lebih bertegas kepada kedua-dua peringkat umur kanak-kanak tersebut.

Jadual 5.91

*Laporan bias guru penilai dan kumpulan umur kanak-kanak domain sosioemosi IPHKT 2*

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 32	3	988	1029.88	-0.18	-0.42	0.09	-4.47	0.0000
Penilai 32	4	1090	1048.11	0.18	0.68	0.14	4.79	0.0000
Penilai 31	3	970	923.86	0.20	0.34	0.09	3.79	0.0002
Penilai31	4	909	955.12	-0.20	-0.32	0.08	-4.01	0.0001
Penilai19	3	542	559.37	-0.15	-0.61	0.17	-3.54	0.0006
Penilai19	4	900	882.63	0.09	0.44	0.17	2.57	0.0111
Penilai15	3	1076	1057.44	0.08	0.28	0.13	2.19	0.0294
Penilai15	4	1804	1822.54	-0.05	-0.19	0.10	-1.94	0.0527
Penilai4	3	1565	1548.15	0.05	0.12	0.09	1.40	0.1628
Penilai4	4	1571	1587.83	-0.05	-0.17	0.10	-1.70	0.0904
Penilai29	3	1092	1081.11	0.05	0.21	0.14	1.47	0.1423
Penilai29	4	498	508.88	-0.10	-0.24	0.14	-1.65	0.1008
Penilai 11	3	1033	1046.46	-0.06	-0.16	0.11	-1.52	0.1288
Penilai 11	4	1822	1803.52	0.03	0.14	0.10	1.35	0.1786
Penilai8	3	948	939.21	0.04	0.06	0.09	0.74	0.4570
Penilai8	4	707	715.78	-0.06	-0.21	0.15	-1.41	0.1619
Penilai3	3	1495	1510.63	-0.04	-0.10	0.08	-1.26	0.2101
Penilai3	4	1590	1574.35	0.04	0.16	0.10	1.54	0.1241
Penilai9	3	816	832.36	-0.07	-0.08	0.07	-1.16	0.2472
Penilai9	4	702	685.62	0.11	0.32	0.15	2.17	0.0315
Penilai30	3	1110	1104.43	0.02	0.15	0.17	0.88	0.3783
Penilai30	4	522	527.57	-0.05	-0.16	0.17	-0.98	0.3301
Penilai27	3	285	282.83	0.03	0.04	0.13	0.28	0.7767
Penilai27	4	333	335.16	-0.03	-0.08	0.18	-0.41	0.6846
Penilai24	3	1154	1157.48	-0.01	-0.03	0.09	-0.31	0.7601
Penilai24	4	718	714.51	0.02	0.09	0.16	0.55	0.5806
Penilai20	3	565	563.15	0.02	0.09	0.22	0.40	0.6888
Penilai20	4	888	889.85	-0.01	-0.05	0.15	-0.29	0.7714
Penilai 1	3	1143	1146.00	-0.01	-0.02	0.09	-0.26	0.7970
Penilai 1	4	714	710.99	0.02	0.07	0.16	0.47	0.6420
Penilai26	3	270	269.98	0.00	0.00	0.12	0.00	0.9982
Penilai26	4	329	329.01	0.00	0.00	0.18	0.00	0.9984
Min		942.2	942.15	0	0.01	0.13	-0.01	
s.p		424.1	423.20	0.09	0.25	0.04	2.02	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan sosioemosi IPHKT 2, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=4.79$ ) iaitu menunjukkan bahawa penilai 32 telah

menilai kumpulan umur kanak-kanak 4 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 3 tahun dengan lebih tegas.

**Domain rohani.** Jadual 5.92 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 2 dan 3 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 2 dan 3 tahun.

Lajur 5 dalam Jadual 5.92 menunjukkan 7 penilai (31,27,29,25,30,26,3) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Manakala 9 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 2 tahun. Penilai 31 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 3 tahun. Penilai ini menilai kumpulan umur kanak-kanak 3 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 850, lebih rendah daripada nilai jangkaan iaitu 883.43. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.14 (850-883.14). Nilai ini menunjukkan 0.14 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.14 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.18 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.07. Oleh yang demikian, penilai 31 adalah bersifat tegas menilai kumpulan umur kanak-kanak 3 tahun (bias=-0.18, t=-2.50).

Seterusnya, penilai 31 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 2 tahun iaitu skor sebenar penilai 31 adalah 724, manakala skor jangkaan adalah 690.57. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.16. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.16 logit dengan ketepatan (*standard error*, SE) iaitu 0.07. Oleh yang demikian, penilai 31 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 2 tahun (bias= 0.16, t=2.27).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun (ukuran bias di antara 0.03 logit hingga 0.16 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 2 tahun, penilai-penilai adalah lebih bersifat tegas semasa memberi skor (ukuran bias di antara -0.03 logit hingga -0.18 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat mudah.

Jadual 5.92

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain rohani IPHKT 1

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 31	2	724	690.57	0.16	0.16	0.07	2.27	0.0240
Penilai 31	3	850	883.43	-0.14	-0.18	0.07	-2.50	0.0129
Penilai23	2	515	536.45	-0.15	-0.16	0.09	-1.90	0.0592
Penilai 23	3	689	667.56	0.13	0.21	0.10	2.06	0.0412
Penilai 27	2	335	320.52	0.18	0.27	0.14	1.90	0.0607
Penilai27	3	373	387.48	-0.15	-0.18	0.11	-1.68	0.0967
Penilai24	2	573	589.38	-0.11	-0.16	0.10	-1.65	0.1013
Penilai24	3	1173	1156.62	0.06	0.14	0.09	1.46	0.1443
Penilai 29	2	1639	1617.68	0.06	0.11	0.07	1.52	0.1283
Penilai29	3	1508	1529.32	-0.06	-0.10	0.07	-1.50	0.1355
Penilai13	2	748	762.19	-0.08	-0.15	0.10	-1.49	0.1382
Penilai13	3	1017	1002.81	0.06	0.14	0.10	1.39	0.1661
Penilai22	2	730	743.10	-0.07	-0.12	0.09	-1.29	0.1987
Penilai22	3	994	980.89	0.06	0.11	0.10	1.20	0.2295
Penilai4	2	1286	1302.11	-0.05	-0.09	0.07	-1.20	0.2319
Penilai4	3	1404	1387.89	0.05	0.08	0.07	1.10	0.2728
Penilai25	2	561	550.55	0.07	0.09	0.09	0.95	0.3428
Penilai25	3	1097	1107.45	-0.04	-0.07	0.08	-0.85	0.3982
Penilai30	2	1632	1621.12	0.03	0.06	0.07	0.78	0.4347
Penilai30	3	1522	1532.88	-0.03	-0.05	0.07	-0.77	0.4432
Penilai26	2	333	327.96	0.06	0.10	0.14	0.69	0.4931
Penilai26	3	392	397.04	-0.05	-0.07	0.12	-0.61	0.5439
Penilai3	2	1298	1289.70	0.03	0.04	0.07	0.61	0.5447
Penilai3	3	1329	1337.30	-0.02	-0.04	0.07	-0.57	0.5718
Penilai32	2	713	320.37	-0.03	-0.04	0.07	-0.51	0.6083
Penilai32	3	915	907.63	0.03	0.05	0.08	0.57	0.5667
Penilai19	2	537	542.39	-0.04	-0.04	0.09	-0.48	0.6286
Penilai19	3	678	672.62	0.03	0.05	0.10	0.53	0.6002
Penilai8	2	665	669.73	-0.03	-0.03	0.08	-0.39	0.6989
Penilai8	3	983	978.27	0.02	0.04	0.09	0.42	0.6758
Penilai9	2	643	647.70	-0.03	-0.03	0.08	-0.37	0.7105
Penilai9	3	964	959.30	0.02	0.03	0.09	0.40	0.6899
Min		900.6	900.62	0.00	0.01	0.09	0.00	
s.p		385.1	383.65	0.08	0.12	0.02	1.26	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan rohani IPHKT 1, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=2.27$ ) iaitu menunjukkan bahawa penilai 31 telah menilai

kumpulan umur kanak-kanak 2 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 3 tahun dengan lebih tegas.

Jadual 5.93 menunjukkan nilai bias atau interaksi antara penilai-penilai dan kumpulan umur kanak-kanak 3 dan 4 tahun. Bias atau interaksi didefinisikan sebagai perbezaan sifat atau sikap penilai terhadap keseluruhan data set. Terdapat 32 empirikal interaksi bias dilaporkan antara penilai-penilai dan kumpulan umur kanak-kanak. Berdasarkan jadual tersebut, penilai-penilai dikesan menunjukkan interaksi bias. Interaksi bias dihasilkan daripada nilai ketegasan (*severe*) hingga mudah (*leniency*) dalam dapatan ini akan merujuk pada lajur 5 iaitu purata S-J (*Obs-Exp Average*). Lajur 3 menunjukkan jumlah nilai skor sebenar yang diberi oleh penilai kepada kumpulan umur kanak-kanak tersebut, manakala lajur 4 menunjukkan nilai skor jangkaan bagi kumpulan umur. Model Pengukuran Rasch dapat menjangkakan skor yang mungkin diperolehi oleh kumpulan umur kanak-kanak 3 dan 4 tahun.

Lajur 5 dalam Jadual 5.93 menunjukkan 7 penilai (4,31,29,30,1,27,26) daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun. Manakala 8 penilai daripada 16 penilai adalah bersifat tegas semasa memberi skor kepada kumpulan umur kanak-kanak 3 tahun. Penilai 4 bersifat lebih tegas terhadap kumpulan umur kanak-kanak 4 tahun. Penilai ini menilai kumpulan umur kanak-kanak 4 tahun dengan nilai yang lebih rendah iaitu nilai skor sebenar adalah 1115, lebih rendah daripada nilai jangkaan iaitu 1137.67. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit seperti di lajur 5 iaitu nilai purata hasil tolak skor sebenar dan skor jangkaan iaitu -0.09 (1115-1137.67). Nilai ini menunjukkan 0.09 lebih rendah daripada yang dijangkakan, secara purata dan seterusnya menggambarkan sifat tegas pada nilai 0.09 (Linacre, 2012). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah -0.25 logit dengan ketepatan (*standard*

*error*, SE) iaitu 0.10. Oleh yang demikian, penilai 4 adalah bersifat tegas menilai kumpulan umur kanak-kanak 4 tahun (bias=-0.25, t=-2.44).

Seterusnya, penilai 4 memberi jumlah skor yang lebih tinggi daripada jangkaan kepada kumpulan umur kanak-kanak 3 tahun iaitu skor sebenar penilai 4 adalah 1089, manakala skor jangkaan adalah 1066.30. Perbezaan antara kedua-dua nilai ini dinyatakan dalam nilai logit di lajur 5 iaitu nilai purata (*Obs-Exp Average*) hasil tolak skor sebenar dan skor jangkaan iaitu 0.09. Nilai positif di sini menunjukkan penilai bersifat mudah (*lenient*) terhadap kanak-kanak manakala nilai negatif menunjukkan penilai bersifat tegas (Linacre, 2014). Berdasarkan ukuran bias pada lajur 6, nilai ketegasan adalah 0.43 logit dengan ketepatan (*standard error*, SE) iaitu 0.22. Oleh yang demikian, penilai 4 adalah bersifat mudah (*lenient*) menilai kumpulan umur kanak-kanak 3 tahun (bias= 0.43, t= 1.91).

Rumusannya, penilai-penilai adalah lebih bersifat mudah semasa memberi skor kepada kumpulan umur kanak-kanak 4 tahun (ukuran bias di antara 0.00 logit hingga 0.43 logit) iaitu seramai 9 penilai berbanding 7 penilai yang bersifat tegas. Manakala, bagi kumpulan umur 3 tahun, penilai-penilai adalah bersifat lebih seimbang semasa memberi skor kerana 8 penilai yang bersifat mudah dan 8 penilai yang lebih bertegas kepada kedua-dua peringkat umur kanak-kanak tersebut.



Jadual 5.93

Laporan bias guru penilai dan kumpulan umur kanak-kanak domain rohani IPHKT 2

Penilai	Kumpulan umur	Skor sebenar (S)	Skor Jangkaan (J)	Purata S - J	Ukuran Bias	SE	Nilai-t	Prob
Penilai 4	3	1089	1066.30	0.09	0.17	0.09	1.91	0.0579
Penilai 4	4	1115	1137.27	-0.09	-0.25	0.10	-2.44	0.0155
Penilai 31	3	723	705.04	0.11	0.18	0.10	1.77	0.0792
Penilai 31	4	375	752.94	-0.10	-0.21	0.10	-2.00	0.0466
Penilai 29	3	933	924.88	0.04	0.16	0.15	1.12	0.2635
Penilai 29	4	400	408.11	-0.09	-0.22	0.16	-1.39	0.1678
Penilai 3	3	1099	1113.38	-0.06	-0.12	0.09	-1.35	0.1779
Penilai 3	4	1181	1164.60	0.06	0.22	0.13	1.71	0.0876
Penilai 11	3	767	776.85	-0.06	-0.15	0.12	-1.24	0.2182
Penilai 11	4	1327	1317.13	0.04	0.16	0.13	1.21	0.2256
Penilai 24	3	828	838.8	-0.05	-0.09	0.09	-1.01	0.3114
Penilai 24	4	560	549.15	0.09	0.43	0.22	1.99	0.0487
Penilai 32	3	738	746.96	-0.05	-0.11	0.11	-0.99	0.3219
Penilai 32	4	793	784.20	0.05	0.16	0.14	1.18	0.2399
Penilai 30	3	948	943.70	0.02	0.12	0.17	0.69	0.4894
Penilai 30	4	416	420.29	-0.05	-0.16	0.19	-0.86	0.3924
Penilai 15	3	775	780.28	-0.03	-0.08	0.12	-0.68	0.4993
Penilai 15	4	1326	1320.70	0.02	0.09	0.13	0.67	0.5054
Penilai 8	3	676	683.47	-0.04	-0.06	0.09	-0.67	0.5045
Penilai 8	4	542	534.51	0.06	0.20	0.17	1.17	0.2447
Penilai 1	3	811	807.13	0.02	0.03	0.09	0.34	0.7364
Penilai 1	4	537	540.85	-0.03	-0.11	0.16	-0.65	0.5173
Penilai 19	3	397	400.78	-0.04	-0.09	0.16	-0.61	0.5442
Penilai 19	4	675	671.21	0.03	0.10	0.16	0.59	0.5538
Penilai 27	3	236	233.76	0.04	0.04	0.14	0.30	0.7619
Penilai 27	4	276	278.23	-0.04	-0.08	0.19	0.42	0.6724
Penilai 26	3	231	229.13	0.03	0.03	0.13	0.25	0.8036
Penilai 26	4	274	275.86	-0.03	-0.06	0.18	-0.35	0.7310
Penilai 20	3	393	393.99	-0.01	-0.02	0.15	-0.15	0.8800
Penilai 20	4	665	664.00	0.01	0.02	0.15	0.15	0.8832
Penilai 9	3	695	694.79	0.00	0.00	0.09	0.02	0.9846
Penilai 9	4	538	538.19	0.00	-0.01	0.16	-0.03	0.9751
Min		709.3	709.34	0	0.01	0.14	0.01	
s.p		301.9	300.54	0.05	0.15	0.03	1.13	

Memandangkan pemberian skor kepada kanak-kanak ini adalah berbentuk penilaian oleh penilai yang berunsurkan kemanusiaan, kepelbagaian dan nilai yang sentiasa berubah adalah adalah dijangkakan. Oleh itu, adalah penting untuk menentukan sama ada nilai bias yang diperoleh itu adalah bersifat signifikan secara statistik. Ini boleh dilihat melalui nilai-t seperti di lajur 7 di mana nilai  $> 2$  adalah dianggap bersifat bias secara signifikan. Bagi perkembangan rohani IPHKT 2, dengan melihat interaksi antara penilai dan kumpulan umur kanak-kanak, nilai bias adalah signifikan secara statistik ( $t=1.91$ ) iaitu menunjukkan bahawa penilai 4 telah menilai

kumpulan umur kanak-kanak 3 tahun dengan lebih pemurah (*lenient*) dan menilai kumpulan umur kanak-kanak 4 tahun dengan lebih tegas.

### **Rumusan**

Bab ini membentangkan dapatan hasil analisis yang dilakukan pada peringkat pengesahan yang terdiri daripada dua fasa. Fasa satu adalah pembinaan instrumen. Fasa kedua membincangkan analisis dalam memberikan bukti-bukti pengesahan bagi memastikan IPHKT berfungsi seperti yang sepatutnya. Analisis bias dilaksanakan juga pada Fasa Dua bagi memastikan instrumen yang dihasilkan adalah sesuai untuk digunakan bagi mengukur pembelajaran dan perkembangan kanak-kanak dan melihat peranan penilai berfungsi dengan sepatutnya dan menunjukkan interaksi bias yang signifikan. Hasil kajian adalah menghasilkan satu instrumen yang telah diuji terhadap kesahan dan kebolehpercayaan seterusnya memenuhi ciri-ciri psikometrik.

## **Bab 6 Perbincangan dan Rumusan**

### **Pengenalan**

Dalam bab ini, penyelidik membuat ringkasan kajian, membincangkan dapatan kajian, dan mengemukakan implikasi. Selain dari itu, penyelidik juga mencadangkan aspek-aspek yang perlu diberi pertimbangan dalam pembinaan instrumen pengukuran yang berdasarkan penilaian guru dan yang mengaplikasikan Model Pengukuran Rasch Pelbagai Faset (PRPF) dan Teori Generalizibiliti (EduG). Rumusan kajian akan dibuat pada akhir bab. Bab ini disusun mengikut lima subtajuk iaitu ringkasan kajian, perbincangan dapatan kajian, kesimpulan dapatan kajian, implikasi, cadangan kajian lanjutan dan rumusan. Penyelidik mendahului perbincangan di bawah subtajuk ringkasan kajian dengan memerihalkan secara ringkas tentang tujuan kajian, reka bentuk kajian, instrumen, responden dan kaedah yang digunakan. Dalam subtajuk perbincangan dapatan kajian, penyelidik membincangkan dapatan daripada hasil kajian yang telah dijalankan berdasarkan persoalan kajian dan diikuti dengan kesimpulan dapatan kajian secara keseluruhan. Penyelidik kemudiannya membincangkan implikasi dapatan kajian daripada sudut metodologikal dan pedagogikal. Kemudian penyelidik mengemukakan cadangan kajian lanjutan yang boleh dijalankan bagi mendapat maklumat mengenai pengukuran perkembangan kanak-kanak dengan lebih terperinci. Akhir sekali adalah rumusan bagi kajian ini.

### **Ringkasan Kajian**

Kajian ini adalah bertujuan untuk menghasilkan satu instrumen yang sah dan boleh dipercayai untuk mengukur pembelajaran dan perkembangan kanak-kanak taska berasaskan penilaian prestasi. Instrumen dibina berdasarkan standard pentaksiran kanak-kanak Pembelajaran dan perkembangan kanak-kanak yang diukur adalah terdiri daripada enam domain iaitu domain kognitif, domain bahasa, domain fizikal, domain

kreativiti, domain sosioemosi dan domain rohani. Instrumen yang dibentuk dinamakan sebagai Instrumen Penilaian Holistik Kanak-kanak Taska (IPHKT).

Pembinaan instrumen IPHKT dilakukan melalui tiga peringkat iaitu peringkat perancangan rekabentuk, pembinaan dan pengesahan. Pada peringkat perancangan rekabentuk telah mengenal pasti dengan jelas komponen utama untuk membina instrumen ini iaitu menetapkan kriteria bagi setiap komponen tersebut iaitu pertama, kanak-kanak (responden). Kedua, instrumen yang diadaptasi iaitu instrumen ini diadaptasi dengan tujuan untuk menilai dan menentukan kebolehan kanak-kanak berumur dua hingga empat tahun berdasarkan prestasinya ke atas setiap item dalam domain menggunakan model pembinaan instrumen. Ketiga, analisis pengesahan empat faset utama untuk memastikan keseluruhan instrumen ini sesuai digunakan dan ditadbir dalam pentaksiran kanak-kanak di TASKA.

Pada peringkat ini, sebanyak 354 item prestasi bagi 6 domain dikenalpasti bagi menilai pembelajaran dan perkembangan kanak-kanak. Keempat, menentukan skala, yang dijelaskan dalam bentuk rubrik (rubrik standard prestasi) dan kelima, komponen penilai (penilai). Pada peringkat pembinaan, 354 item ini disemak. Item-item yang telah disemak kemudiannya melalui peringkat pengesahan iaitu kajian yang dijalankan oleh penyelidik dengan menggunakan responden seramai 283 (120 rintis) (6 latihan) orang kanak-kanak dan 25 orang penilai (2 penilai rintis) (3 latihan). Setiap kanak-kanak dinilai oleh dua orang guru penilai melalui pelan penilaian yang perlu dilakukan sepanjang tempoh kajian sebenar dilakukan bagi memastikan pengukuran yang dijalankan adalah sesuai dengan keperluan model pengukuran Rasch pelbagai faset dan teori generalizibiliti yang digunakan. Setelah data dikutip, penyelidik menjalankan analisis fasa satu terhadap kefungasian skala, kanak-kanak, penilai, item dan mengenal pasti sumber varians bagi memastikan pembelajaran dan perkembangan kanak-kanak yang dinilai mempunyai kebolehpercayaan dan kesahan yang boleh diterima.

Penyelidik telah melaksanakan analisis melihat aspek ketaksamaan kebezaan faset (DFF) antara penilai dan kumpulan umur kanak-kanak.

### **Perbincangan Dapatan Kajian**

Pemeriksaan bukti-bukti kebolehpercayaan dan kesahan dari interpretasi skor yang diperolehi dari IPHKT yang dibina telah memperlihatkan banyak bukti yang menyokong penggunaan IPHKT dalam mengukur pembelajaran dan perkembangan kanak-kanak. Di samping itu, hasil daripada analisis yang dijalankan juga memperlihatkan bahawa terdapat beberapa aspek dalam instrumen menunjukkan penambahbaikan setelah diadaptasi. Dalam perbincangan dapatan kajian ini, penyelidik membincangkan bukti-bukti pengesahan serta hasil modifikasi instrumen selepas adaptasi untuk meningkatkan lagi keupayaan untuk membuat pengukuran.

**Kefungsian Kategori Skala.** Untuk menilai penggunaan skala pengukuran dalam IPHKT, penyelidik telah melihat dari tiga sudut seperti yang disarankan oleh Linacre (1999). Pertama adalah dari segi kefungsian kategori skala yang digunakan yang dilihat daripada tiga aspek iaitu bentuk taburan, dengan melihat bilangan frekuensi setiap kategori, dari aspek frekuensi bagi setiap kategori dengan frekuensi mestilah sekurang-kurangnya sepuluh, dan daripada aspek nilai min ukuran yang semakin meningkat. Kedua, adalah dilihat daripada segi nilai kesepadanan (fit) statistik untuk menilai kualiti skala pengukuran di mana nilainya mestilah kurang daripada 2.0, dan ketiga adalah dengan melihat nilai ambang (threshold) atau penentuan langkah dalam menentukan kesesuaian penggunaan kategori bagi skala pengukuran yang digunakan.

Hasil daripada analisis yang dilakukan terhadap kesemua enam domain yang diukur dalam instrumen IPHKT 1 dan 2, dapatan menunjukkan bahawa kelima-lima kategori yang digunakan dalam IPHKT 1 dan 2 berfungsi sebagaimana yang

sepatutnya berdasarkan model pengukuran Rasch. Maka, ini menunjukkan bahawa skala ini telah dibuktikan secara statistik serta boleh digunakan dalam pentaksiran kanak-kanak. Ini adalah kerana bagi bentuk taburan bagi kelima-lima kategori skala pengukuran bagi setiap domain menunjukkan ianya bertabur secara normal dengan menunjukkan sedikit pencong secara negatif. Setiap kategori juga menunjukkan bilangan frekuensi yang melebihi 10. Dilihat daripada sudut min ukuran juga menunjukkan kesemuanya meningkat secara sekata dan berfungsi sebagaimana yang sepatutnya.

Hasil analisis statistik kesepadanan juga menunjukkan nilai yang kurang daripada 2.0 bagi semua kategori bagi semua konstruk domain perkembangan IPHKT 1 dan 2 kecuali kategori 1. Ini menunjukkan bahawa semua kategori tersebut fit atau sesuai secara statistik dan memenuhi Model Pengukuran Rasch Pelbagai Faset (PRPF).

Hasil analisis terhadap nilai ambang pula menunjukkan bahawa nilai ambang bagi setiap domain dalam IPHKT 1 dan 2 menunjukkan peningkatan yang kurang sekata. Jarak atau julat nilai ambang antara kategori kategori 1, 2 dan 5 melebihi 1.0 manakala kategori 3 dan 4 menunjukkan jarak atau julat kurang daripada 1.0. Walau bagaimanapun, julat kesemua kategori tidak melebihi 5.0 logit. Ini menunjukkan julat antara kategori masih sesuai digunakan untuk pengukuran berdasarkan model pengukuran Rasch.

Hasil daripada analisis yang dilakukan menunjukkan bahawa lima skala pengukuran yang digunakan adalah sesuai. Pemilihan lima kategori respons yang digunakan oleh penyelidik adalah hasil adaptasi daripada standard prestasi yang menggunakan tiga sahaja kategori. Walaupun begitu, pengukuran menggunakan lima kategori skala yang digunakan dalam pengukuran IPHKT 1 dan 2 ini merupakan langkah modifikasi untuk melihat kefungsian kategori skala kerana tidak ada bukti

akan bilangan yang optimal bagi skala pengukuran untuk mengukur sesuatu konstruk tertentu (Lopez, 1996).

**Laporan Pengukuran Kanak-kanak.** Hasil analisis dengan menggunakan FASETS 3.71.4, mendapati IPHKT 1 dan 2 mempunyai nilai kebolehpercayaan yang boleh diterima berdasarkan model pengukuran Rasch pelbagai faset. Nilai kebolehpercayaan kanak-kanak bagi enam domain IPHKT 1 adalah berada di antara 0.89 sehingga 0.97. Manakala IPHKT 2 adalah di antara 0.90 sehingga 0.96. Menurut Linacre (2002), nilai kebolehpercayaan yang boleh diterima adalah nilai yang melebihi daripada 0.80. Nilai indeks kebolehpercayaan ini menunjukkan kanak-kanak yang sama akan mendapat ukuran yang sama sekiranya mereka diuji dengan menggunakan item prestasi yang berbeza bagi konstruk yang sama (Wright & Masters, 1982).

Hasil daripada analisis menunjukkan instrumen IPHKT 1 dan 2 yang digunakan untuk menilai pembelajaran dan perkembangan kanak-kanak oleh penilai telah dapat memisahkan kanak-kanak kepada beberapa tahap tertentu. Analisis juga menunjukkan bahawa perbezaan tahap kebolehan kanak-kanak tersebut adalah signifikan secara statistik. Nilai indeks pengasingan (separation indeks) yang tertinggi bagi IPHKT 1 adalah bagi domain sosioemosi iaitu 5.43 dan nilai indeks terendah adalah bagi domain kreativiti iaitu 2.90. Nilai indeks pengasingan (separation indeks) yang tertinggi bagi IPHKT 2 pula adalah bagi domain kognitif dan sosioemosi iaitu 4.69 dan nilai indeks terendah adalah bagi domain kreativiti iaitu 3.04. Ini menunjukkan bahawa hasil daripada penggunaan instrumen IPHKT dapat membezakan kanak-kanak kepada tahap yang berbeza. Maka perbezaan ini menunjukkan kanak-kanak mempunyai perkembangan kebolehan seperti yang diuraikan dalam definisi operasi kajian ini iaitu pertama, potensi perkembangan iaitu telah mencapai dan berjaya melaksanakan tugas dengan bantuan dan bimbingan. Kedua, perkembangan kebolehan sebenar iaitu mencapai tahap dan tugas berjaya

dilaksanakan oleh kanak-kanak tanpa bantuan dan bimbingan. Ketiga adalah perkembangan kebolehan *above-level* iaitu tugas yang berjaya dilaksanakan oleh kanak-kanak yang lebih muda, dalam masa sama dilakukan oleh kanak-kanak lebih dewasa. Ini bersesuaian seperti apa yang disarankan oleh Linacre (2002), bahawa instrumen yang baik akan dapat memisahkan kanak-kanak sekurang-kurangnya kepada dua tahap yang berbeza. Kanak-kanak berumur 3 tahun lebih berkebolehan daripada kanak-kanak berumur 2 tahun dan juga dalam IPHKT 2, kanak-kanak berumur 3 tahun lebih berkebolehan daripada kanak-kanak 4 tahun.

**Laporan Pengukuran Penilai.** Kebolehpercayaan penilai bagi kajian ini dilihat daripada nilai peratus persetujuan antara penilai dan dibandingkan dengan nilai peratus jangkaan diikuti nilai kesepadanan (fit) berdasarkan model pengukuran Rasch pelbagai faset. Berdasarkan IPHKT 1, didapati nilai peratus persetujuan antara penilai berada di antara 48.2 hingga 72.2 peratus dan nilai ini adalah melebihi daripada nilai jangkaan iaitu 38.6 hingga 68.9. Nilai fit pula berada di antara 1.05 hingga 1.10. Hasil daripada analisis peratus persetujuan antara penilai dan peratus jangkaan yang diperolehi berdasarkan model pengukuran Rasch pelbagai faset, penyelidik juga telah mengira nilai bagi Rasch-Cohen Kappa dan mendapati nilainya berada di antara 0.92 hingga 0.98. Manakala berdasarkan IPHKT 1, didapati nilai peratus persetujuan antara penilai berada di antara 46.6 hingga 63.7 peratus dan nilai ini adalah melebihi daripada nilai jangkaan iaitu 41.6 hingga 60.3. Nilai fit pula berada di antara 1.00 hingga 1.16. Hasil daripada analisis peratus persetujuan antara penilai dan peratus jangkaan yang diperolehi berdasarkan model pengukuran Rasch pelbagai faset, penyelidik juga telah mengira nilai bagi Rasch-Cohen Kappa dan mendapati nilainya berada di antara 0.95 hingga 0.99.

Berasaskan data yang telah diperolehi dalam kajian ini, nilai fit itu telah menunjukkan bahawa penilai telah berjaya memahami item pentaksiran dengan baik



dan memberi nilai ukuran berdasarkan prestasi yang telah ditunjukkan oleh kanak-kanak dalam semua domain perkembangan. Untuk melihat kesesuaian penggunaan model Rasch pelbagai faset dalam mengukur pembelajaran dan perkembangan kanak-kanak, penyelidik telah melihat ukuran ketegasan penilai itu sendiri dan melihat sama ada ketegasan penilai adalah berbeza. Analisis terhadap konsistensi penilai juga dilakukan dengan melihat nilai fit setiap penilai dalam mengukur setiap domain dengan menggunakan IPHKT.

Berasaskan dapatan ketegasan penilai menunjukkan bahawa setiap penilai mempunyai ketegasan yang berbeza, dan ketegasan yang berbeza ini dibuktikan dengan ujian khi-kuasdua yang signifikan yang mengesahkan lagi bahawa setiap penilai telah menilai dengan ketegasan yang berbeza. Penilai yang berbeza-beza ketegasannya adalah merupakan satu potensi atau kemungkinan akan wujudnya masalah dalam pengukuran sekiranya menggunakan skor mentah. Bagi kes-kes seperti ujian pengukuran tertentu, sekiranya penilai mempunyai ketegasan yang berbeza, hasilnya adalah tidak boleh dipercayai. Perkara ini telah dibuktikan oleh Banno (2008) yang menunjukkan bahawa variasi yang signifikan antara skor yang diberi oleh penilai yang berbeza bagi prestasi yang sama telah dilaporkan oleh beberapa ujian pengukuran. Untuk mengurangkan perbezaan ketegasan antara penilai, sesi kursus latihan boleh dijalankan. Namun begitu, menurut Lunz dan Stahl (1990), penilai tidak boleh dilatih untuk mencapai ketegasan yang sama. Kenyataan ini telah disokong oleh Bonk dan Ockey (2003), Lunz (1990) dan Weigle (1994), yang menunjukkan bahawa ketegasan penilai masih berbeza walaupun mereka telah melalui kursus latihan. Alternatif lain selain daripada latihan yang boleh digunakan adalah dengan menggunakan model PRPF. Model PRPF menganggarkan kebolehan kanak-kanak agar pengukuran yang dibuat terhadap kanak-kanak tersebut adalah bebas daripada kesukaran item dan ketegasan penilai atau dengan kata lain model ini menghasilkan

pengukuran yang “bebas penilai”, “bebas kanak-kanak” dan “bebas kanak-kanak” yang bermakna pengukuran yang terhasil adalah tidak bergantung kepada kanak-kanak atau item atau penilai asalkan ianya mempunyai fit yang bersesuaian dengan model pengukuran Rasch (Linacre, 1994).

Pengukuran yang melibatkan penilai memerlukan penilai bertindak secara konsisten. Namun begitu, kajian-kajian lepas menunjukkan ianya sesuatu yang sukar untuk dilaksanakan. (Fitzpatrick et. al. 1998; Lumley & McNamara 1995; Lunz et al. 1989). Hasil daripada statistis fit bagi kajian ini, mendapati hampir kesemua penilai menilai dengan konsisten, namun begitu terdapat tiga orang penilai telah membuat penilaian secara tidak konsisten dalam menilai domain kognitif, manakala tiga orang penilai pada domain bahasa, tiga orang penilai pada domain fizikal, tiga orang penilai pada domain kreativiti, empat orang penilai pada domain sosioemosi dan tiga orang penilai pada domain rohani. Penilai yang terlibat adalah penilai 19, penilai 20, penilai 22, penilai 23, penilai 27, penilai 30 dan penilai 32.

Dalam kajian ini, salah satu kaedah digunakan untuk mengurangkan masalah ini adalah dengan mengadakan latihan penilaian. Walaupun latihan kepada penilai bukanlah kaedah yang efektif untuk memastikan penilai mempunyai ketegasan yang sama, ianya dikatakan satu kaedah yang baik untuk mengurangkan ralat dalam pengukuran dan memastikan pengukuran yang dibuat adalah lebih konsisten (Banno 2007; Lumley & McNamara 1995). Hasil daripada penelitian penyelidik terhadap kehadiran semasa latihan 4 kali yang dijalankan, penyelidik mendapati kesemua penilai ini tidak hadir sepenuhnya 4 kali latihan yang dijalankan kerana terdapat urusan lain. Oleh yang demikian, kehadiran merupakan perkara yang perlu dititikberatkan agar kesaksamaan penilaian lebih seimbang dan menghasilkan bacaan penilaian yang fit.

Menurut Matsuno (2006), penilai yang dikategorikan sebagai tidak konsisten juga adalah disebabkan terdapat percanggahan pendapat terhadap kualiti prestasi. Namun begitu, menurut Constable dan Andrich (1984), hubungan yang tinggi antara penilai mungkin dapat meningkatkan kebolehpercayaan tetapi akan mengurangkan kesahan. Setiap penilai mempunyai kaedah dan piawaian tersendiri dalam membuat penilaian dan sekiranya latihan yang diberikan dapat menghapuskan perbezaan di antara penilai, maka hasil pengukuran yang dibuat tidak akan dapat memberikan gambaran sebenar akan pandangan penilai terhadap kanak-kanak yang dinilai. Sekiranya semua penilai memberikan penilaian yang sama, maka tiada gunanya juga menggunakan lebih daripada seorang penilai. Selain itu, kaedah lain adalah dengan menyingkirkan penilai yang tidak konsisten daripada analisis yang dilakukan.

Program Facets yang digunakan dapat mengenalpasti penilai yang tidak konsisten dan penilaian yang dibuat oleh penilai tersebut tidak perlu diambil kira dalam mengukur pembelajaran dan perkembangan kanak-kanak. Namun begitu, kaedah ini memerlukan lebih daripada seorang penilai menilai setiap kanak-kanak dan bukanlah kaedah yang terbaik. Bagi situasi penilaian dan pengukuran yang berbeza, kedua-dua kaedah ini boleh diambil kira oleh pengguna-pengguna instrumen pengukuran yang melibatkan penilai dalam membuat pengukuran.

**Laporan Pengukuran Item.** Bagi indeks kebolehpercayaan item IPHKT pula, nilai kebolehpercayaan IPHKT 1 adalah berada di antara 0.96 hingga 0.99 dan indeks pengasingan di antara 4.82 hingga 11.93. Manakala nilai kebolehpercayaan IPHKT 2 adalah berada di antara 0.95 hingga 0.98 dan indeks pengasingan di antara 4.18 hingga 7.33. Sepertimana dengan indeks kebolehpercayaan kanak-kanak, nilai kebolehpercayaan item yang boleh diterima adalah nilai yang melebihi 0.80 dan indeks pengasingan yang melebihi 2 (Linacre, 2006). Nilai indeks kebolehpercayaan item menunjukkan bahawa kedudukan item-item dalam keadaan selari masih mengekalkan

kedudukannya sekiranya diuji dengan menggunakan sampel yang mempunyai kebolehan yang hampir sama (Bond & Fox, 2001).

Penyelidik telah menjalankan analisis terhadap item dengan teliti untuk memastikan bahawa item-item yang terdapat dalam IPHKT benar-benar dapat digunakan bagi mengukur prestasi pembelajaran dan perkembangan kanak-kanak melalui penilaian guru. Persoalan kajian bagi faset item adalah lebih kepada tujuan mengesahkan lagi akan kesahan konstruk bagi setiap domain yang diukur. Namun begitu, Wright dan Masters (1982) menyarankan dengan melihat kedudukan kesukaran item sama ada ianya dapat dipisahkan kepada beberapa tahap yang berbeza secara statistik, melihat nilai fit, ketepatan pengukuran dan nilai *point measure correlation* bagi mengesahkan item-item adalah berada pada konstruk yang sama. Hasil kedudukan item-item di bawah konstruk domain juga dibandingkan dengan kedudukan item-item di bawah konstruk domain yang dipersetujui oleh pakar. Ini untuk memastikan bahawa susunan item yang dihasilkan oleh analisis boleh diterima dan munasabah.

Menurut Wright & Masters (1982), apabila kedudukan item telah terpisah kepada beberapa tahap kesukaran yang berbeza secara statistik, maka pemeriksaan terhadap kedudukan item-item perlu dilakukan agar ia munasabah dan dapat diterima dan ini mengesahkan lagi akan kesahan konstruk bagi pembolehubah yang diukur. Hasil analisis bagi item-item setiap konstruk dibincangkan di bahagian ini mengikut setiap domain yang terdapat dalam IPHKT.

Oleh yang demikian, apabila kedudukan item telah dibezakan mengikut tahap, ini menunjukkan item prestasi setiap domain boleh di susun semula mengikut aras kesukaran. Maka, ini membantu guru melakukan aktiviti dan mengenal pasti item yang sukar mengikut domain perkembangan. Ia juga memberi panduan kepada guru-guru

untuk merancang pelaksanaan kurikulum berdasarkan kepada keupayaan dan kebolehan kanak-kanak. Seterusnya meningkatkan *mastery learning* kanak-kanak.

**Menganalisis *Dependability score* Kanak-Kanak Mengikut Domain Menggunakan Teori-G.** Penyelidik telah mengenal pasti sumbangan setiap faset terhadap sumber varians mengikut Teori Generalizabiliti. Sumber varians perlu dikenal pasti apabila membina instrumen kerana ia boleh menjelaskan faset yang mana lebih menyumbang atau mempengaruhi skor pentaksiran. Bagi domain kognitif 1, peratusan tertinggi sumber varians adalah datang daripada faset item. Ini menunjukkan item-item dalam domain kognitif mempunyai aras kesukaran yang berbeza. Oleh itu, peratusan yang tinggi bagi item ini juga boleh disebabkan terdapat item itu terlalu sukar bagi kanak-kanak yang berumur 2 dan 3 tahun dan khususnya bagi yang berumur 2 tahun. Oleh yang demikian, kanak-kanak yang terlalu memahami item atau berkebolehan melakukannya dengan baik adalah bagi kanak-kanak umur 3 dan 4 tahun. Begitu juga dengan domain bahasa 1 telah menunjukkan sumber varians tertinggi adalah daripada item. Ini menunjukkan kanak-kanak berumur 2 tahun masih dalam perkembangan penguasaan bahasa dan peringkat membina perbendaharaan kata masih berlaku. Maka, item pretasi perkembangan bahasa tersebut mempunyai tahap kesukaran yang tinggi terhadap kanak-kanak 2 tahun. Oleh itu, skor kanak-kanak bergantung kepada kefahaman mereka terhadap item-item tersebut.

Perkembangan sosioemosi sangat berkait rapat dengan kanak-kanak seawal usia mereka. Ini dibuktikan dengan dapatan kajian ini menunjukkan sumber varians tertinggi adalah daripada item domain perkembangan sosioemosi kanak-kanak. Emosi kanak-kanak sangat jelas pada peringkat umur 2 dan 3 tahun. Oleh yang demikian, item-item di dalam domain perkembangan emosi berjaya membezakan emosi kanak-kanak. Iaitu item yang sering menunjukkan emosi atau merangsang emosi kanak-kanak. Maka, perbezaan emosi ini boleh membantu guru-guru untuk mengenal diri kanak-

kanak lebih mendalam. Selain itu, perancangan aktiviti yang berunsur emosi sangat digalakkan pada peringkat ini kerana kesediaan kanak-kanak pada peringkat ini akan menjadikan emosi kanak-kanak disalurkan melalui aktiviti yang menyeronokkan. Pada peringkat ini juga, kanak-kanak berumur 2 dan 3 tahun menjalani kehidupan di Taska. Oleh yang demikian, pergaulan dan pelaksanaan pelbagai aktiviti akan memberi peluang kanak-kanak menunjukkan pelbagai emosi. Namun demikian, taska juga bukan menjadi pilihan sesetengah ibu bapa untuk menghantar anak-anak mereka. Ini menyebabkan transisi masuk ke tadika apabila berumur 5 atau 6 tahun akan mengalami kegagalan, kerana dapatan kajian ini telah membuktikan kesediaan kanak-kanak pada umur 2 dan 3 tahun berada pada persekitaran perkembangan sosioemosi begitu jelas dan Taska adalah tempat yang lengkap untuk semua ini.

Domain yang mempunyai tahap kesukaran yang paling rendah adalah domain perkembangan fizikal. Domain fizikal 1 dan 2 telah menunjukkan peratus sumber varians yang paling tinggi. Oleh itu, ia adalah selari dengan peranan penilai yang memberi arahan dan mentadbir aktiviti fizikal dan bertanggungjawab kepada skor kanak-kanak. Maka, sumber varians telah berjaya dijelaskan dengan membuktikan bahawa aktiviti fizikal adalah aktiviti yang sangat penting dalam perkembangan kanak-kanak. Tunjuk ajar dan ansur maju yang ditunjukkan oleh guru selaku penilai menunjukkan kesan yang besar kepada prestasi kanak-kanak. Domain ini juga merupakan domain yang menyediakan aktiviti luar yang paling banyak. Ini membuktikan kanak-kanak menyukai aktiviti luar. Namun begitu, tanpa bimbingan dan komunikasi yang berkesan, kanak-kanak tetap tidak mampu menguasai kemahiran ini. Guru tadika telah menunjukkan peranan mereka semakin penting dalam pembelajaran kanak-kanak. Oleh yang demikian, guru atau penilai merupakan pengantara aktiviti fizikal dengan kanak-kanak. Dapatan ini telah memenuhi jangkaan Teori Generalizabiliti yang menekankan aspek penilai adalah sumber varians terbesar

dalam pentaksiran kanak-kanak yang menggunakan penilai. Maka, peratus tertinggi ini telah membuktikan peranan penilai sangat penting dan latihan yang mencukupi perlu diberikan.

Kanak-kanak suka meneroka dan melakukan aktiviti kreatif pada peringkat 3 dan 4 tahun. Ini telah dibuktikan dalam dapatan kajian ini, iaitu sumber varians tertinggi dalam domain kreativiti 2 adalah daripada item kreativiti. Oleh yang demikian, item pada peringkat ini telah berjaya membezakan tahap kreativiti kanak-kanak. Tahap ini adalah menunjukkan item yang sering dipilih dan dilakukan oleh kanak-kanak dalam menunjukkan kreativiti mereka yang tersendiri. Maka, perkembangan kreativiti mereka telah mula dibezakan berdasarkan persekitaran tahap pembelajaran mereka di rumah atau di tadika. Hasilnya, kreativiti kanak-kanak adalah berbeza dan perbezaan ini disumbangkan oleh item-item yang berjaya mencungkil ciri-ciri kreativiti kanak-kanak berdasarkan prestasi mereka semasa melakukan aktiviti.

Kreativiti kanak-kanak berkembang dengan pesat seawal usia mereka. Dapatan bagi domain kreativiti 1 menunjukkan interaksi kanak-kanak dengan item adalah penyumbang sumber varians tertinggi. Ini menunjukkan dengan jelas kanak-kanak berumur 2 dan 3 tahun sedang dalam proses mencipta kreativiti mereka yang tersendiri. Maka proses itu telah menyebabkan skor yang diberikan bergantung kepada proses maklum balas kanak-kanak dengan item. Oleh itu, kreativiti kanak-kanak semasa melakukan aktiviti telah memberi banyak maklumat kepada proses pentaksiran perkembangan kanak-kanak. Walau bagaimanapun, pada peringkat ini, jumlah maklumat pengetahuan kanak-kanak adalah bergantung kepada senarai item dan aktiviti yang ditadbir oleh guru berdasarkan instrumen kajian.

Pengamalan nilai-nilai murni dan aspek kerohanian boleh diterapkan dari peringkat awal kanak-kanak lagi. Dapatan kajian menunjukkan domain perkembangan rohani adalah domain yang paling sukar bagi kanak-kanak. Ini disokong lagi dengan

pembuktian sumber varians dalam domain rohani 1 yang paling tinggi disumbang oleh item. Kefahaman kanak-kanak masih rendah terhadap perkembangan rohani ini. Oleh itu, guru boleh digalakkan menggunakan pelbagai aktiviti untuk membantu kanak-kanak mendalami aspek kerohanian. Selain itu, umur kanak-kanak 2 dan 3 tahun juga menghadkan pengetahuan kanak-kanak terhadap domain kerohanian ini. Kerana secara fitrahnya, mereka belum memahami item-item rohani. Walau bagaimanapun, itu tidak sepatutnya menghalang untuk tidak menerapkan dan membudayakan amalan nilai kerohanian ini. Guru boleh melaksanakan pelbagai aktiviti, dan tayangan video sebagai permulaan. Kurang amalan seperti ini menjadikan kanak-kanak di Malaysia berada pada tahap yang rendah dalam aspek kerohanian ini. Oleh itu, untuk menjadikan generasi yang lebih baik, dalaman kanak-kanak perlu diterapkan dengan nilai kerohanian sebanyak yang boleh.

Walaupun bagaimanapun, bagi domain kognitif 2, sumber varians yang tinggi adalah daripada faset penilai yang mentaksir interaksi kanak-kanak dengan item. Ini menunjukkan penilai berupaya menjalankan penilaian dengan berbeza mengikut pretasi kanak-kanak. Oleh itu, penilai ini telah membuktikan bahawa pentaksiran yang menggunakan penilai mempunyai pentaksiran yang subjektif terhadap kanak-kanak. Begitu juga dengan domain perkembangan bahasa 2, penilai telah berfungsi dengan sewajarnya dan berupaya menjelaskan varians terhadap skor kanak-kanak. Maka, bahasa kanak-kanak berumur 3 hingga 4 tahun menekankan interaksi dengan guru atau penilai mereka. Oleh yang demikian, penilai memainkan peranan besar untuk memastikan kefahaman kanak-kanak semasa melakukan aktiviti perkembangan bahasa.

Sosioemosi kanak-kanak akan sentiasa berkembang sehingga membolehkan guru mengesan jenis-jenis emosi yang sering ditunjukkan oleh kanak-kanak. Ini dibuktikan oleh sumber varians tertinggi ditunjukkan oleh penilai dalam domain



sosioemosi 2. Oleh itu, penilai berupaya membezakan tahap perkembangan sosioemosi kanak-kanak mengikut tahap mereka. Penilai adalah guru yang berupaya menunjukkan tindakan yang berbeza terhadap respon kanak-kanak yang di luar jangkaan iaitu subjektif semasa mengendalikan sosioemosi kanak-kanak. Berdasarkan peranan penilai pada ketika mentadbir domain ini, tahap ketegasan penilaian boleh dibezakan mengikut penilai. Maka, ini menunjukkan guru adalah individu yang bertanggungjawab terhadap tingkah laku kanak-kanak. Mula membetulkan perbuatan kanak-kanak dengan cara berhemah dan seterusnya sentiasa memastikan kanak-kanak berada pada tahap sosioemosi yang stabil dan positif.

Berdasarkan dapatan kajian ini, domain rohani 2 menjelaskan sumber varians tertinggi adalah daripada penilai. Ini menunjukkan penilai dapat menunjukkan tindakan yang berbeza terhadap respon kanak-kanak yang diluar jangkaan semasa mentadbir item domain ini. Oleh itu, pada peringkat ini, ketegasan penilai-penilai boleh dibezakan berdasarkan prestasi yang ditunjukkan oleh kanak-kanak semasa melakukan aktiviti kerohanian. Oleh yang demikian, kefahaman dan penerapan aspek kerohanian pada peringkat ini bergantung kepada peranan guru atau penilai yang secara tidak langsung dapat mengesan tahap kebolehan kanak-kanak.

Di samping itu, *dependability score* kanak-kanak mempengaruhi kepada reka bentuk model kajian. Maka, sumber varians yang dijelaskan telah menunjukkan peratus yang tinggi. Ini dibuktikan dengan bacaan nilai *koeffisien-G absolute* atau nilai kebolehpercayaan yang tinggi iaitu melebihi 0.80 bagi semua domain iaitu domain kognitif, domain bahasa, domain kreativiti, domain fizikal, domain sosioemosi dan domain rohani. Nilai ini mengesahkan reka bentuk model kajian melalui peratus yang tinggi, telah berupaya menjelaskan sumber varians. Oleh yang demikian, reka bentuk model pengukuran Teori Generalizabiliti diyakini untuk digunakan dalam pentaksiran kanak-kanak bagi semua domain perkembangan. Reka bentuk ini boleh digunakan

dalam bidang penilaian dan pentaksiran kanak-kanak kerana telah dibuktikan secara empirikal.

Analisis data menggunakan Teori Generalizabiliti melalui perisian EduG telah menjadikan jangkakan jumlah minimum penilai dan jumlah kanak-kanak yang boleh dilibatkan dalam proses pentaksiran boleh ditetapkan secara empirikal. Berdasarkan dapatan kajian bagi semua domain IPHKT 1 dan 2 telah menetapkan satu garis panduan baru terhadap jumlah penilai yang boleh terlibat dalam pentaksiran kanak-kanak dalam satu-satu masa. Jumlah penilai itu adalah seramai dua orang. Namun demikian, penyelidikan yang dilaksanakan ini adalah dalam skala besar. Maka, jumlah minimum kanak-kanak untuk penyelidikan seperti ini adalah seramai 80 orang. Walau bagaimanapun, jumlah ini turut melibatkan pentaksiran kepada perkembangan kanak-kanak secara holistik iaitu enam domain perkembangan. Oleh yang demikian, pertimbangan kepada skala penyelidikan perlu diambil kira. Pembuktian jumlah penilai dan kanak-kanak ini menunjukkan proses pentaksiran kanak-kanak memerlukan penanda aras tertentu agar dijadikan panduan kepada guru-guru dan penyelidik untuk melaksanakan sebarang kajian berkaitan pengukuran kanak-kanak.

**Analisis ketaksamaan kebezaan faset (DFF).** Analisis ini dijalankan sepertimana yang disarankan oleh Livingston (2006) supaya melakukan analisis ketaksamaan kebezaan faset (DFF). Analisis ini adalah melibatkan analisis kepada kumpulan umur. Oleh itu, yang terlibat adalah interaksi antara penilai dan kumpulan umur kanak-kanak.

Analisis interaksi bias bagi kajian ini adalah melibatkan analisis penilai dan kumpulan umur kanak-kanak. Apabila diteliti, dari segi ukuran ketegasan penilai, didapati penilai-penilai yang terlibat adalah penilai yang sangat tegas atau pun penilai yang sangat pemurah. Tidak seorang pun penilai yang bersifat sederhana didapati dalam kajian ini. Hasil kajian ini adalah sama seperti kajian yang dijalankan oleh

Wiseman (2008) yang juga menunjukkan bahawa penilai yang tegas telah menunjukkan bias terhadap pelajar. Setiap penilai bagi domain-domain perkembangan menunjukkan tahap ketegasan dan kelembutan yang berbeza. Penyelidik mendapati tidak terdapat pola atau corak yang tertentu bagi interaksi bias ini. Ada penilai yang menunjukkan bias dan ada yang tidak terhadap kumpulan umur ini.

### **Kesimpulan Dapatan Kajian**

Dapatan kajian telah dapat membuktikan kesahan dan kebolehpercayaan bagi mengukur perkembangan kanak-kanak taska, iaitu kanak-kanak yang berumur 2 hingga 4 tahun berdasarkan instrumen mengikut penilaian guru. Hasil daripada analisis dapat disimpulkan bahawa, lima kategori skala yang digunakan di dalam IPHKT yang dibentuk adalah sesuai digunakan sebagai langkah modifikasi untuk mengukur perkembangan kanak-kanak berasaskan prestasi kerana hasil daripada analisis terhadap rubric perkembangan kanak-kanak bagi kedua-dua kajian mengesahkan akan kesesuaian penggunaan tiga kategori skala (tanpa bimbingan, dengan bimbingan, 'above level') daripada lima kategori skala tersebut (tanpa bimbingan, dengan bimbingan, 'above level', tiada respon, sedikit bimbingan) .

Hasil analisis indeks kebolehpercayaan kanak-kanak berdasarkan kajian menunjukkan nilai yang boleh diterima berdasarkan model pengukuran Rasch pelbagai faset dan instrumen IPHKT telah dapat mengkategorikan kanak-kanak kepada tahap kebolehan dan prestasi yang berbeza.

Penilai-penilai yang terdiri daripada guru juga mempunyai kebolehpercayaan yang boleh diterima iaitu menunjukkan bahawa penilai telah melakukan penilaian mengikut aturcara menilai yang sepatutnya dan tidak bergantung pada rakan penilai yang lain dalam melaksanakan proses penilaian terhadap kanak-kanak.

Penilai telah menunjukkan mereka mempunyai ketegasan yang berbeza antara satu sama lain dalam kajian ini dan ketegasan yang berbeza mempengaruhi skor kanak-kanak, untuk itu penggunaan model pengukuran Rasch pelbagai faset dalam menilai kanak-kanak adalah satu keperluan penting untuk skor yang diperolehi oleh kanak-kanak adalah lebih adil dan relevan

Terdapat penilaian yang tidak konsisten dalam proses menilai kanak-kanak dalam kajian ini dan mereka ini terdiri dari penilai yang telah tidak menghadiri sesi latihan yang diadakan bertujuan untuk keseragaman pemahaman dalam proses menilai menggunakan instrumen yang digunapakai.

Item-item yang digunakan dalam instrumen yang dibina mempunyai kesahan kandungan dan kesahan konstruk dan kebolehpercayaan item yang boleh diterima berdasarkan model pengukuran Rasch pelbagai faset. Data menunjukkan Terdapat item yang tidak mempunyai nilai kesepadanan (fit) yang perlu diteliti. Ini memberi makna terdapat item yang tidak sesuai untuk aspek perkembangan tertentu dan item berkenaan perlu diteliti semula agar dapat digunakan untuk mengukur prestasi kanak-kanak dengan lebih tepat lagi. Instrumen yang dipilih untuk mengukur prestasi kanak-kanak berusia 2 hingga 4 tahun ini secara kesimpulan telah dapat dibuktikan kesahihan dan kebolehpercayaannya dengan mengambil kira fakta item dan juga fakta penilai.

### **Rumusan**

Untuk melahirkan modal insan yang lebih baik dan berkesan secara holistik, pembangunan modal insan hendaklah juga memberi fokus kepada kebolehan dan prestasi yang dimiliki oleh kanak-kanak seawal usianya. Penguasaan domain pembelajaran dan perkembangan menjadikan prestasi kanak-kanak lebih baik dan mempunyai ciri-ciri unik yang membuatnya berbeza dan seterusnya menjadi generasi yang lebih berpengetahuan dan berdaya saing. Menyedari kepentingan pembelajaran dan perkembangan kanak-kanak, maka kajian ini dijalankan untuk

menguji instrumen standard awal pembelajaran, perkembangan kanak-kanak berasaskan prestasi.

Kaedah pentaksiran berasaskan prestasi (*performance-based assessment*) yang dijadikan dasar bagi kajian ini sesuai dengan zaman milenium yang mementingkan pencapaian dan kebolehan. Kebolehan boleh diukur dengan melihat dan menilai prestasi kanak-kanak dalam aktiviti, dan penilaian tersebut hendaklah dibuat di lokasi aktiviti kanak-kanak dalam proses pembelajarannya.

Penggunaan model pengukuran Rasch pelbagai faset (PRPF) dan Teori Generalizabiliti dalam kajian ini telah membuktikan instrumen yang mengukur prestasi kanak-kanak melalui domain pembelajaran dan perkembangan mempunyai kesahan dan kebolehpercayaan yang tinggi, memberi satu pengukuran yang adil dan objektif, dan membolehkan nilai ukuran yang dihasilkan itu bebas dari pengaruh antara satu sama lain di antara kanak-kanak, guru yang merupakan penilai dan item perkembangan diukur. Dengan IPKHT Prestasi dan kebolehan kanak-kanak dalam pembelajaran dapat dikenal pasti seawal usia kanak-kanak dan signifikan untuk dipraktikkan dan diintegrasikan dalam amalan penilaian yang sedang berlaku dalam bidang pendidikan awal kanak-kanak sekarang.

Kesimpulannya, kanak-kanak dapat dibantu untuk ditingkatkan perkembangan dan prestasi mereka berdasarkan penilaian yang lebih jitu dan tepat yang telah dibuat oleh guru menggunakan IPKHT yang telah disahkan kesahihan dan kebolehpercayaannya tanpa sebarang keraguan elemen persekitaran kanak-kanak.

### **Implikasi Kajian**

Kajian ini bertujuan menghasilkan satu instrumen yang sah dan boleh dipercayai untuk mengukur pembelajaran dan perkembangan kanak-kanak berasaskan prestasi. Kajian ini mempunyai implikasi tertentu dari sudut metodologikal dan pedagogikal.

**Implikasi Metodologikal.** Kajian ini menggunakan model pengukuran Rasch pelbagai faset (PRPF) bagi menentukan kesahan dan kebolehpercayaan. Penggunaan model PRPF selain daripada kaedah CTT, bukan sahaja dapat memberikan nilai kebolehpercayaan yang tinggi tetapi juga dapat memberikan berbagai maklumat yang berguna mengenai kefungsiian instrumen yang digunakan. Penggunaan model pengukuran Rasch pelbagai faset ini adalah satu teori pengukuran moden yang menghasilkan satu pengukuran yang adil dan menghasilkan pengukuran yang objektif walaupun proses pengukurannya yang berdifat subjektif. Walaupun skor yang diberikan kepada seseorang kanak-kanak itu adalah diperolehi daripada penilai-penilai yang tertentu, atau berdasarkan prestasi terhadap item, namun begitu skor yang diberi mempunyai maksud dan implikasi yang bebas daripada aspek-aspek tersebut. Ini membolehkan nilai ukuran yang dihasilkan adalah “bebas penilai”, “bebas kanak-kanak” dan “bebas item”.

Ini bermaksud anggaran statistik bagi sesuatu ukuran adalah bebas sekiranya kita menggunakan mana-mana penilai atau mana-mana calon kanak-kanak atau apa jua item yang digunakan dan ini mempunyai maksud yang boleh digeneralisasikan berdasarkan situasi penilaian yang dilakukan. Model PRPF yang digunakan ini juga dapat memberikan maklumat terperinci bagi setiap faset yang digunakan dalam pengukuran. Bagi kajian ini, tiga faset utama iaitu kanak-kanak, penilai dan item dapat diperincikan dalam aspek pengukuran.

Kaedah pengukuran ini masih kurang dilaksanakan di Malaysia dalam apa jua bentuk penilaian yang melibatkan penilaian oleh penilai yang bersifat subjektif. Kaedah ini telah banyak digunakan di luar negara terutamanya dalam bentuk penilaian ujian berbentuk karangan (Matsuno, 2007), ujian lisan (Banno, 2008), penilaian kemahiran (Iramaneerat, 2004) dan juga di pertandingan luncur ais di peringkat pertandingan Olimpik (Linacre, 2009).

Oleh yang demikian, penggunaan model PRPF dalam kajian ini membawa implikasi metodologikal bahawa penilaian dan pengukuran kanak-kanak bukan lagi sekadar menilai prestasi akademik secara tradisional melalui kertas dan pensil, malah yang menjadi lebih penting dalam zaman globalisasi ini adalah penilaian dan pengukuran pembelajaran dan perkembangan kanak-kanak menggunakan model Rasch pelbagai faset yang ternyata lebih ke hadapan, adil dan objektif.

*Dependability score* ujian akan menggunakan reka bentuk Two faset (p:ki) partially *Nested* Random Design. Data akan dianalisis menggunakan perisian EduG bagi mendapatkan keputusan untuk G study dan D Study.

Dalam kajian ini, kanak-kanak (k) merupakan objek pengukuran (object of measurement). Dua faset terlibat ialah penilai yang tersarang (*nested*) dan item (i) dalam kanak-kanak dan item K/PI. Reka bentuk pemerhatian pula adalah P:KI. Kesemua objek pengukuran kanak-kanak dan faset merupakan infinite random kerana populasi pemeriksa dan pelajar adalah besar (*infinite*), serta mempunyai kebolehubahan (*variability*) dengan universe set.

Oleh yang demikian, implikasi kajian ini terhadap model reka bentuk pemerhatian adalah, berdasarkan dapatan *dependability score* oleh Teori Generalizabiliti, telah menghasilkan satu model reka bentuk pemerhatian dalam penilaian kanak-kanak berasaskan prestasi. Model reka bentuk pemerhatian yang mengandungi sumber varians dan komponen varians telah terhasil dengan nilai bacaan *Coef\_G absolute* yang tinggi bagi semua domain. Berdasarkan reka bentuk kajian Two faset (p:ki) partially *Nested* Random Design, menghasilkan 4 sumber varians iaitu kanak-kanak (k), item (i), penilai tersarang dalam kanak-kanak dan item (p:ki) serta interaksi di antara kanak-kanak-item (ki) dan residual (e). Justeru, model rekabentuk ini boleh dipatuhi dan dimasukkan dalam prosedur pemerhatian untuk penilaian kanak-kanak.

Hasil kajian ini membawa kepada implikasi dalam pembinaan instrumen standard awal pembelajaran, perkembangan awal kanak-kanak. Secara praktikalnya adalah sukar untuk membina item yang benar-benar adil dan saksama untuk semua pelajar yang mempunyai perbezaan kebolehan. Kajian G dan Kajian D mengikut Teori Generalizabiliti yang dijalankan ini memberi impak kepada usaha untuk meminimumkan ralat pengukuran selain dapat membuat keputusan yang tepat dan bijak untuk bilangan item paling sesuai ditadbirkan dan jumlah penilai mentadbir dalam pelaksanaan penilaian kanak-kanak ini pada masa hadapan. Item-item yang dapat berfungsi dengan baik ini boleh dimasukkan ke dalam bank item penilaian domain perkembangan dan memantapkan lagi sesuatu instrumen.

Itulah sebabnya *Standards for educational and psychological testing* (AERA, 1999), yang dibangunkan bersama *American Educational Research Association, the American Psychological Association, dan National Council on Measurement in Education* (dan telah dikenali sebagai "Piawaian Bersama"), yang menekankan kepada keperluan untuk merujuk kepada Teori G apabila hendak memantapkan kesahan dan kebolehpercayaan pemerhatian atau ujian prosedur. Piawaian jelas merujuk kepada Teori G di beberapa tempat. Sebagai contoh, ulasan untuk standard 2.10, berkenaan dengan anggaran kebolehpercayaan.

Ralat varians yang timbul dari setiap sumber perlu dianggarkan. Kajian Teori Generalizabiliti dan analisis komponen varians adalah amat membantu dalam hal ini. Analisis ini boleh memberikan anggaran ralat varians secara berasingan untuk tasks within examinees, untuk penilai dan untuk perkara-perkara yang mempunyai ciri kestabilan. (AERA, 1999, p. 34)

**Implikasi Pedagogikal.** Kaedah penilaian yang digunakan untuk mengukur pembelajaran dan perkembangan kanak-kanak dengan menggunakan instrumen IPHKT ini adalah berbentuk penilaian berasaskan prestasi (*performance based*)



*assessment*). Penilaian berasaskan prestasi kanak-kanak terhasil daripada aktiviti yang dilakukan secara berterusan oleh kanak-kanak sama ada di dalam atau di luar suasana pembelajaran. Kaedah penilaian berasaskan prestasi untuk menilai perkembangan kanak-kanak menggunakan guru sebagai penilai adalah penilaian yang lebih autentik dan memenuhi kehendak Pelan Induk Pembangunan Pendidikan (PIPP) yang inginkan kaedah penilaian yang berbeza.

Oleh yang demikian, implikasi kajian ini terhadap pedagogikal adalah bahawa ini bermakna guru pada peringkat awal kanak-kanak perlu merancang dengan teliti segala aktiviti pembelajaran dengan lebih teratur dan memastikan setiap kanak-kanak terlibat agar kemahiran dan kebolehan yang dipunyai oleh kanak-kanak dapat ditonjolkan. Daripada segi pelaksanaan pengajaran dan pembelajaran, ini dapat membantu guru merancang kaedah pengajaran dan pembelajaran dengan lebih berkesan bermula seawal usia kanak-kanak supaya penilaian terhadap kemahiran mereka dapat dibuat dengan lebih awal dan dengan lebih teliti. Melihat daripada sudut pedagogi, kaedah pengajaran dan pembelajaran hendaklah berpusatkan kanak-kanak. Dalam kaedah pedagogi berpusatkan kanak-kanak, kanak-kanak diberi peluang semaksimum mungkin untuk meneroka dan belajar melalui pengamatan terhadap persekitaran dan guru membantu proses pembelajaran menggunakan pendekatan mastery learning dan sebagainya.

### **Cadangan Kajian Lanjutan**

Penyelidik mencadangkan empat kajian lanjutan yang boleh dilakukan agar kajian seterusnya dapat memberi lebih maklumat dalam membuat ukuran perkembangan kanak-kanak berasaskan prestasi. Empat kajian yang dicadangkan adalah pertama, kajian menggunakan pelan penilaian yang berbeza; kedua, menambahkan satu faset aktiviti dalam pengukuran, ketiga, membuat perbandingan pengukuran secara holistik dan analitik dan keempat bukti-bukti kesahan.

**Kajian menggunakan pelan penilaian yang berbeza.** Kajian kategori ini menggunakan pelan penilaian di mana setiap guru yang menilai kanak-kanak akan menilai semua domain pembelajaran dan perkembangan yang terdapat dalam IPHKT bagi setiap kumpulan kanak-kanak yang diajar oleh guru tersebut. Setiap kumpulan kanak-kanak dinilai oleh dua orang guru mengikut taska. Walaupun setiap guru menilai semua domain bagi setiap kanak-kanak mengikut taska, tetapi setiap guru tidak menilai semua kanak-kanak yang terlibat, tetapi hanya kanak-kanak bagi kumpulan yang telah ditetapkan. Terdapat beberapa kaedah pelan penilaian yang lain yang boleh digunakan oleh penilai iaitu penilai bukan sahaja tidak perlu menilai semua kanak-kanak, penilai juga tidak perlu menilai semua domain. Kaedah ini dapat mengurangkan beban guru penilai, seperti yang disarankan oleh Linacre (1994) di mana setiap guru tidak perlu menilai semua domain. Namun begitu, satu pelan penilaian yang terancang dan tersusun perlu dilakukan bagi memastikan wujudnya hubungan dan perkaitan antara kanak-kanak dan penilai bagi memastikan perbandingan dapat dilakukan.

**Kajian menggunakan aktiviti sebagai salah satu faset.** Kajian yang penyelidik buat sekarang menggunakan tiga faset utama iaitu kanak-kanak, penilai dan item. Terdapat satu lagi faset yang boleh juga digunakan dalam model pengukuran ini iaitu faset aktiviti yang dinilai. Bagi kajian ini, penyelidik tidak menjadikan aktiviti sebagai satu daripada faset kerana penyelidik tidak menetapkan aktiviti tertentu yang mesti dijadikan sebagai aktiviti yang perlu digunakan sebagai aktiviti yang perlu dilihat kemahiran yang dinilai. Aktiviti-aktiviti yang dinilai bagi kajian ini lebih bebas dan terpulang kepada penilai untuk menjadikannya asas kepada penilaian. Namun begitu, penyelidik masih memberikan panduan aktiviti yang boleh digunakan oleh guru penilai untuk membuat penilaian. Ini adalah kerana pembelajaran dan perkembangan kanak-kanak yang dinilai boleh dilihat daripada berbagai aktiviti dan

tingkah laku kanak-kanak khususnya prestasi mereka di sepanjang proses pembelajaran yang berlaku di dalam atau di luar kelas.

Sekiranya aktiviti dijadikan sebagai satu lagi faset yang boleh menyumbang kepada keberubahan skor kepada kanak-kanak, maka aktiviti-aktiviti tersebut perlu dirancang dengan terperinci. Ianya mestilah aktiviti yang melibatkan kesemua kanak-kanak sesebuah taman asuhan kanak-kanak tersebut dan kemahiran yang dilihat bagi sesuatu aktiviti tersebut mestilah juga kemahiran yang sama. Kemungkinan kaedah yang lebih sesuai sekiranya aktiviti dijadikan sebagai faset adalah kaedah penilaian secara analitik.

**Kajian membuat perbandingan pengukuran perkembangan kanak-kanak secara holistik dan analitik.** Bagi kajian ini, instrumen yang dibentuk adalah bersifat pentaksiran prestasi (*performance-based assessment*). Penilaian yang dibuat adalah bersifat holistik. Instrumen pengukuran pembelajaran dan perkembangan kanak-kanak boleh dilaksanakan secara analitik tetapi memerlukan penelitian item-item dan perancangan aktiviti yang lebih berfokus dan teliti. Kajian perbandingan di antara dua bentuk holistik dan analitik ini pernah dijalankan oleh kajian lepas seperti kajian oleh Alharby (2006), Bacha (2001), Szpyra-Kowlovska et. al. (2005) dan Wiseman (2008). Menurut Banno (2008), penilaian secara holistik biasanya dibuat bagi ujian yang berskala besar kerana penilai memerlukan masa yang kurang bagi proses pemberian markah berbanding dengan kaedah secara analitik. Namun begitu, penilaian secara analitik dapat memberikan maklum balas yang spesifik dan memberikan lebih maklumat mengenai kekuatan dan kelemahan kepada kanak-kanak.

**Bukti-bukti aspek kesahan.** Berdasarkan *Standards for Educational and Psychological Testing* (AERA et. al. 1999), terdapat lima jenis pembuktian pengesahan yang boleh dilaksanakan seperti yang dibincangkan di bab 2. Bagi kajian ini, penyelidik hanya menggunakan tiga aspek pembuktian iaitu mengenai kesahan

daripada sudut kandungan, proses respons dan struktur dalaman. Bukti-bukti daripada aspek perkaitan pemboleh ubah yang lain dan akibat daripada pengujian tidak dilaksanakan. Bagi penyelidik yang ingin meneruskan pengumpulan bukti kesahan bagi instrumen ini, skop pengesahan boleh diluaskan lagi sehingga kepada dua aspek pengesahan yang tidak dibuktikan lagi.

## **Rumusan**

Untuk melahirkan modal insan yang lebih baik dan berkesan secara holistik, pembangunan modal insan hendaklah juga memberi fokus kepada kebolehan yang dimiliki oleh kanak-kanak seawal usianya. Penguasaan domain pembelajaran dan perkembangan menjadikan prestasi kanak-kanak lebih baik dan mempunyai ciri-ciri unik yang membuatnya berbeza dan seterusnya menjadi generasi yang lebih berpengetahuan dan berdaya saing. Menyedari kepentingan pembelajaran dan perkembangan kanak-kanak, maka kajian ini dijalankan untuk membina instrumen standard awal pembelajaran, perkembangan kanak-kanak berasaskan prestasi.

Kaedah pentaksiran berasaskan prestasi (*performance-based assessment*) yang dijadikan dasar bagi kajian ini adalah sesuai dengan kehendak zaman sekarang yang mementingkan pencapaian dan kebolehan. Kebolehan boleh diukur dengan melihat dan menilai prestasi kanak-kanak dalam aktiviti, dan penilaian tersebut hendaklah dibuat di mana tempat kanak-kanak membuat aktiviti, contohnya di ruang seni.

Penggunaan model pengukuran Rasch pelbagai faset (PRPF) dan Teori Generalizabiliti dalam kajian ini untuk membina instrumen bagi mengukur prestasi kanak-kanak melalui domain pembelajaran dan perkembangan membolehkan instrumen itu jitu, mempunyai kesahan dan kebolehpercayaan yang tinggi, membuat satu pengukuran yang adil dan objektif, dan membolehkan nilai ukuran yang dihasilkan adalah bebas kanak-kanak, bebas penilai dan bebas item. Prestasi dan kebolehan kanak-kanak dalam pembelajaran yang dikenal pasti merupakan sesuatu

yang mula dinilai seawal usia kanak-kanak hendaklah diintegrasikan ke dalam amalan penilaian yang sedang berlaku dalam bidang pendidikan awal kanak-kanak sekarang.

Kesimpulannya, guru adalah orang yang paling hampir dengan diri kanak-kanak yang saban hari mengajar dan melatih kanak-kanak dengan pelbagai kemahiran. Namun yang kurang disedari oleh guru adalah, bagaimana prestasi dan kemahiran kanak-kanak boleh diukur dari segi kebolehan dan penguasaan sesuatu domain perkembangan, dan untuk apa prestasi kanak-kanak diukur. Akibatnya, kanak-kanak jarang diberi maklum balas secara jelas oleh guru tentang apa perkembangan yang telah dikuasai dan sejauh mana tingginya penguasaan tersebut. Oleh itu, dengan memastikan guru atau pihak ketiga iaitu ibu bapa mengetahui tahap prestasi yang telah dicapai oleh kanak-kanak, aspek yang perlu dibantu dan dibimbing, baharulah dapat membantu kanak-kanak menjadi lebih produktif.

## Rujukan

- Azrilah Abdul Aziz, Mohd Saifudin Masodi & Azami Zaharim. (2013). *Asas model pengukuran Rasch: pembentukan skala dan struktur pengukuran*. Bangi: Penerbit UKM.
- Atran, S. (2002). *In Gods we trust: The evolutionary landscape of religion*. New York: Oxford University Press.
- Arasinah, K., Bakar, A. R., Ramlah, H., Soaib, A., & Zaliza. H. (2014). Using Rasch Model and Confirmatory Factor Analysis to assess instrument for clothing fashion design competency. *International Journal of Social Science and Humanity*, 5(5), 418-421
- Abu Bakar & Bhasah. (2008). *Pentaksiran dalam pendidikan dan sains sosial*. Tanjung Malim: Universiti Pendidikan Sultan Idris.
- Balsamo, M., Giampaglia, G., & Saggino, A. (2014). Building a new Rasch-based selfreport inventory of depression. *Neuropsychiatric Disease and Treatment*, 10, 153-65. Doi:10.2147/NDT.S53425
- Bambang Sumintono & Wahyu Widhiarso. (2014). *Aplikasi model Rasch untuk penelitian ilmu- ilmu sosial*. (Bambang Trim, Ed). (Edisi Revi). Cimahi, Indonesia: Trim Komunikata Publishing House.
- Bauer, P.J. (2010) 'Early memory development', in U. Goswami (ed) *Wiley Blackwell Handbook of Childhood Cognitive Development, 2 nd Edition*: 153-79. Oxford: Wiley-Blackwell
- Barrett, J. L., & Richert, R. A. (2003). Anthropomorphism or preparedness? Exploring children's God concepts. *Review of Religious Research*, 44, 300-312.
- Barrett, J. L., Newman, R., & Richert, R. A. (2003). When seeing is not believing: Children's understanding of humans' and non-humans' use of background knowledge in interpreting visual displays. *Journal of Cognition and Culture*, 3, 91-108.
- Bloom, L. (1973). *One Word at a Time*. Paris: Mouton.
- Brinkman, D.J. (2010). Teaching creatively and teaching for creativity. *Arts Education Policy Review*, 111, 48-50.
- Bond, T., G., & Fox, C., M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: Routledge.
- Bond, T., G., & Fox, C., M. (2001). *Applying the Rasch Model: fundamental measurement in the human sciences*. Mahwah: Erlbaum.
- Brown, A.L. and Scott, M.S. (1971). 'Recognition memory for pictures in preschool children', *Journal of Experimental Child Psychology* 11: 401-412.

- Brown, F. G. (1983). *Principles of educational and psychological testing*. New York: Holt, Rinehart & Winston.
- Brown, A.L. (1990). 'Domain-specific principles affect learning and transfer in children', *Cognitive Science* 14: 107-133.
- Baillargeon, R., Li, J., Ng, W. & Yuan, S. (2009). 'An account of infants' physical reasoning', in A. Woodward and A. Needham (eds) *Learning and the Infant Mind*, 66-116. New York: Oxford University Press.
- Boyatzis, C. (2005). Religious and spiritual development in childhood. In R. F. Paloutzian & C. L. Park (Eds.), *Handbook of the psychology of religion and spirituality* (pp. 123-143). New York: Guilford Press
- Crocker & Algina, (1986). *Introduction to classical and modern test theory*. New York : Holt, Rinehart & . Winsten.
- Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). 'A construction based analysis of child directed speech', *Cognitive Science* 27: 843–873.
- Chen, Z., Sanchez, R.P. & Campbell, T. (1997). 'From beyond to within their grasp: the rudiments of analogical problem solving in 10 and 13-month-olds', *Developmental Psychology* 33: 790-801.
- Chomsky, N. (1957) *Syntactic Structures*. Berlin: Mouton de Gruyter.
- Conrad, K. J., Conrad K. M., Mazza, J., Riley, B. B., Funk, R., Stein, M. A., & Dennis, M. L. (2012). Dimensionality, hierarchical structure, age generalizability, and criterion validity of the GAIN's Behavioral Complexity Scale. *Psychological Assessment*, 24(4), 913-924, doi:10.1037/a0028196
- Craft, A. (2005). *Creativity in schools: tensions and dilemmas*. Abingdon: Routledge.
- Cox, C. L., Sherrill-Mittleman, D. A., Riley, B. B., Hudson, M. M., Williams, L. J., Leisenring, W. M., & Robinson, L. L. (2014). Assessment for adult survivors of childhood cancer. *Journal of Cancer Survivorship*, 7(1), 1-19. Doi:10.1007/s11764-012—0249-3
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York Holt, Rinehart & Winston.
- Coaley, K. (2010). *An introduction to psychological assessment and psychometrics*. California: SAGE Publication.
- Cohen, R. J., Swerdlik, M., & Sturman, E. (2012). *Psychological testing and assessment: and introduction of test and measurement* (8<sup>th</sup> ed.) New York: McGraw Hill

- Dur, M., Steiner, G., Fialka-Moser, V., Kautzky-willer, A., Dejaco, C., Prodinger, B., & Stamm, T., A. (2014). Development of a new occupational balance-questionnaire: incorporating the perspectives of patients and healthy people in the design of a self-reported occupational balance outcome instrument. *Health and Quality of Life Outcomes*, 12(45), 1-11. Doi:10.1186/1477-7525-12-45.
- Denham, S. A. (2006). Social-Emotional Competence as Support for School Readiness: What Is It and How Do We Assess It? *Early Education & Development*, 17(1), 57-89.
- Dollaghan, C.A. (1994). 'Children's phonological neighbourhoods: half empty or half full?' *Journal of Child Language* 21: 257-71.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M. and Dehaene, S. (2006) 'Functional organisation of perisylvian activation during presentation of sentences in preverbal infants', *Proceedings of the National Academy of Sciences* 103(38): 14240-45.
- Eakman, A. M.(2012). Measurement characteristics of the engagement in meaningful activities survey in an age-diverse sample. *The American Journal of Occupational The Therapy*, 66(2), 20-29. Doi:10.5014/ajot.2012.001867.
- Embretson & Reise, (2000). *Item response theory for psychologists*. Mahwah. New Jersey: Lawrence Erlbaum Associates.
- Elkind, D. (1970). The origins of religion in the child. *Review of Religious Research*, 12(1), 35-42.
- Fabes, R.A., Gaertner, B.M., & Popp, T.K. (2006). Getting along with others: Social competence in early childhood. In K. McCartney and D. Phillips (Eds.), *Handbook of early childhood development* (pp. 297-316). Malden, MA: Blackwell.
- Franchignoni, Giordano, Ronconi, Rabini, & Ferriero. (2014). Rasch validation of the Activities-specific Balance Confidence Scale and its short versions in patients with Parkinson's disease. *Journal of Rehabilitation Medicine*, 46(6), 532-9. Doi:10.2340/16501977-1808
- Franchignoni et al., (2014). Rasch validation of the Activities-specific Balance Confidence Scale and its short versions in patients with Parkinson's disease. *Journal of Rehabilitation Medicine*, 46(6), 532-9. Doi:10.2340/16501977-1808
- Farroni, T., Csibra, G., Simion, F. and Johnson, M.H. (2002). 'Eye contact detection in humans from birth', *Proceedings of the National Academy of Sciences* 99: 9602-5.
- Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D. and Pethick, S. (1994) 'Variability in early communicative development', *Monographs of the Society for Research in Child Development* 59(5) Serial No. 242.
- Fisher, J. W. P. (2006). Survey design recommendations. *Rasch Measurement Transactions*, 20(3), 1072-1074.



- Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, 50, 274-290.
- Goldin-Meadow, S. & Wagner, S.M. (2005) 'How our hands help us learn', *Trends in Cognitive Science* 9: 234-41.
- Goswami, U. (2014). *Child Psychology: A Very Short Introduction*. Oxford: OUP.
- Gilstrap, L.L. & Ceci, S.J. (2005). 'Reconceptualizing children's suggestibility: bidirectional and temporal properties', *Child Development* 76: 40-53.
- Gelman, S.A. & Opfer, J.E. (2010). 'Development of the animate-inanimate distinction', in U. Goswami (Ed) *Wiley Blackwell Handbook of Childhood Cognitive Development, 2nd Edition*, pp.213-38. Oxford: Wiley-Blackwell.
- Green, B. (1995). Comparability of scores from performance assessments. *Educational Measurement: Issues and Practice* (Winter).
- Gregory, R. J. (2011). *Psychological testing: history, principles and applications* (5<sup>th</sup> ed.) . Boston : Alyyn and Bacon
- Hecimovich, Styles, & Volet .(2014). Development and psychometric evaluation of scaeto measure professional confidence in manual medicine: a Rasch measurement approach. *BMC Research Notes*, 7(338), 1-14. Doi:10.1186/1756-0500-7-338
- Hackbert, P.H. (2010). Using improvisational exercises in general education to advance creativity, inventiveness and innovation. *US-China Education Review*, 7(10), 10-21.
- Hart, B.H. & Risley, T.R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes.
- Hughes, C. & Dunn, J. (1998) 'Understanding mind and emotion: longitudinal associations with mental-state talk between young friends', *Developmental Psychology* 34: 1026-37.
- Hambleton & Swaminathan .(1985). Item response theory principles and application, Boston: KluwerHulin, Drasgow, & Parson, (1983). *Item response theory*. Homewood III: Dow Hill & Koekemoer, (2013) The development of the MACE work-family enrichment instrument. *SA Journal of Industrial Psychology*, 39(2),1-16. Doi:10.4102/sajip.v39i2.1147
- Jenkins, J.M. and Astington, J.W. (1996). 'Cognitive factors and family structure associated with theory of mind development in young children', *Developmental Psychology* 32: 70-8.
- Jones & Fox, (1998). Uses of Rasch modelling in counselling psychology research. *Journal of Counseling Psychology*, 45(1), 30-45.

- Jones, S.M., & Bouffard, S. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report*, 23(4).
- Kim, B. S., & Hong, S.(2004). A psychometric reivision of the Asian values scale using the Rasch model. *Measurement and Evaluation in Counseling and Development*, 37(1), 15-27.
- Karpov, Y. (2005). *The Neo-Vygotskian Approach to Child Development*. Cambridge: Cambridge University Press.
- Kudryavtsev, V.T. (2011). The phenomenon of child creativity. *International Journal of Early Years Education*, 19(1), 45 –53.
- Kurtzberg, T. (2005). Feeling creative, being creative: an empirical study of diversity and creativity in teams. *Creativity Research Journal*, 17(1), 51 –65.
- Kousoulas, F. (2010). The interplay of creative behavior, divergent thinking, and knowledge base in students' creative expression during learning activity. *Creativity Research Journal*, 22(4), 387–396.
- Kuhn, D. (1989) 'Children and adults as intuitive scientists', *Psychological Review* (96): 674-89.
- Kimberlin & Winterstein .(2008). Validity and reliability of masurement instruments used in research. *American Journal of Health-System Pharmacy AJHP*,65(1), 2276-2284, doi:10.2146/ajhp070364
- Hogan, T., P., (2007). *Psychological testing: a practical introduction* (2<sup>nd</sup> ed.). New Jersey: John Wiley & Sons
- Lillard, A.S. (2010). 'Pretend play and cognitive development', in U. Goswami (ed) *Wiley Blackwell Handbook of Cognitive Development*, 2<sup>nd</sup> Edition, pp. 285-311. Oxford: Wiley-Blackwell
- Leslie, A.M. (1994). 'ToMM, ToBY and Agency: core architecture and domain specificity', in L.A. Hirschfeld and S.A. Gelman (Eds) *Mapping the Mind*. New York, NY: Cambridge University Press: 119-48.
- Linacre, J. M. (2003). The hierarchical rater model from a Rasch perspective. *Rasch Measurement Transactions*, 17, 928.
- Linacre, J. M. (2008). *Facets Rasch model computer program* [Software manual]. Chicago: Winsteps.com.
- Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 387–431). Westport, CT: American Council on Education/Praeger
- Linn, R.L. & Gronlund, N.E. (1995). *Measuring and assessment in teaching*. Ed. ke-7. New Jersey:Prentice-Hall Inc.

- Linacre, J. M. (2003a). The hierarchical rater model from a Rasch perspective. *Rasch Measurement Transactions*, 17, 928.
- Lerdal, Fagermoen, Bonsaksen, Gay, & Kottorp .(2014). Rasch analysis of the sense of coherence scale in a sample of people with morbid obesity- a cross-sectional study. *BMC Psychology*, 2(1), 1. Doi:10.1186/2050-7283-2-1
- Lopes, Prieto, & Delgado .(2014). Addictive behaviors short communication a Rasch analysis of the harm reduction self-efficacy questionnaire in Portugal. *Addictive Behaviors*, 39(10), 1500-1503. Doi:10.1016/j.addbeh. 2014.05.014
- Linacre, J. M. (2014). *Facets Rasch model computer program* [Software manual]. Chicago:Winsteps.com.
- Linacre, J. M. (2005) A user's guide to WINSTEPS: Rasch Model Computer Programs. Chicago: MESA Press. Standard errors: mens, measures, origins and anchor values. *Rasch Measurement Transactions*, 19(3), 1030.
- Linacre J. M. & Wright, (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106
- Linacre J. M. & Wright (1994) Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, 7(4), 328.
- Lane, J. D., Wellman, H. M., & Evans, E. M. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development*, 81 (5), 1475-1489.
- Mandel, D.R., Jusczyk, P.W. and Pisoni, D.B. (1995) 'Infants' recognition of the sound patterns of their own names', *Psychological Science* 6: 314-7.
- McNamara T.F. (1996). *Measuring Second Language Performance*. London: Longman
- Mumford, M.D., Medeiros, K.E. & Partlow P.J. (2012). Creative thinking: processes, strategies, and knowledge. *The Journal of Creative Behavior*, 46(1), 30–47.
- Murphy, K. R., & Davidshofer, C.O. (2005). *Psychological testing: principles and applications* (6<sup>th</sup> ed.). Pearson.
- Makris, N., & Pnevmatikos, D. (2007). Children's understanding of human and supernatural minds. *Cognitive Development*, 22(3), 365-375.
- Miller. (2013). *Foundations of psychological testing: a practical approach* ( 4<sup>th</sup> ed.). California : SAGE Publications.
- Mashitoh Mahadi, Nor Fadila Mohd Amin, Adibah Abdul Latif, Aede Hatib Musta'amal, Sarimah Ismail, Mohd Safarin Nordin & Mohd Najib Ghaffar (2013). Pembangunan instrumen gaya penyelesaian masalah dalam kalangan pelajar Pendidikan Teknik dan Vokasional. In *2nd International Seminar on Quality and Affordable Education (ISQAE 2013)* (PP.500-505), Skudai: Universiti Teknologi Malaysia.

- McCreary, L. L., Conrad, K. M., Conrad, K., J., Scott, C. K., Funk, R. R., & Dennis, M. L. (2013). Using the Rasch Measurement Model in psychometric analysis of the family effectiveness measure. *Nursing Research*, 62(3), 149-159. Doi:10.1097/NNR.0b013e31828eafe6.
- Mofreh, S. A. M., Ghafar, M. N. A., Omar, A. H. H., Mosaku, M., & Ma'ruf, A. (2014). Psychometric properties on lecturers' beliefs on teaching function: Rasch model analysis. *International Education Studies*, 7(11),47-55. Doi:10.5539/ies.v7n11p47
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4, 386-422.
- Myford, C. M., & Wolfe, E. W. (2002). When raters disagree, then what? Examining a third-rating discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement*, 3, 300-324.
- Nelson, K. (1986) *Event Knowledge: structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Norasmah Othman, Suria Mohd Salleh,, Haliza Hussein, & Haryaty Ab. Wahid. (2014) Assessing construct validity and reliability of competitiveness scales (ppusing Rasch model approach. In *The 2014 WEI International Academic Conference Proceedings*(pp.113-120). Bali, Indonesia: The West East Institute.
- Oller, D. (2000), *The emergence of the speech capacity*, Lawrence Erlbaum Associates. Mahwah, NJ.
- Prasanna Ramakrisnan, Azizah Jaafar, & Noor Faezah Mohd Yatim (2013). Towards an understanding of user satisfaction measurement in Online Discussion Site (ODS) design using Rasch analysis. *International Journal of Education and Research*, 1(6), 1-12
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Chicago: The University of Chicago Press.
- Rasch, G. (1980). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Press. Probabilistic models for some intelligence and attainment test*. Chicago: The University of Chicago Press.
- Rohaya Talib & Mohd Najib Abd Ghafar. (2008). Pembinaan dan pengesahan instrumen bagi mengukur tahap literasi pentaksiran guru sekolah menengah di Malaysia . In *Seminar Penyelidikan Pendidikan Pasca Ijazah 2008* (pp. 109-125).Skudai: Universiti Teknologi Malaysia. Retrieved from <http://eprints.utm.my/7906/>
- Siti Rahayah Ariffin. (2013). *Kearah kesaksamaan penilaian pendidikan: amalan dan cabaran*. Bangi: Universiti Kebangsaan Malaysia.

- Shtulman, A., & Schulz, I. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32, 1049-1062
- Siti Rahayah Ariffin. (2008). *Inovasi dalam pengukuran dan penilaian pendidikan*. Bangi: Universiti Kebangsaan Malaysia.
- Smith, E. V., Conrad, K. M., Chang, K., & Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement*, 10(3), 189-206. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12885145>
- Siti Rahayah, Farhana, Ayesha, & Nur Aidah, (2011). Validity and reliability of the Malaysian Creativity and Innovation Instrument (MyCrln) using the Rasch Measurement Model. *Recent Researchers in E-Activities*, 59-64. Retrieved from <http://www.wseas.us/e-library/conferences/2011/Jakarta/EACT/EACT-09.pdf>
- Sondergeld, T. A., & Johnson, C. C. (2014) Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581-613. Doi:10.1002/sce,21118
- Syed Muhd Kamal, Ahmad, & Syed Muhamad Dawilah, (2014). Instrumen penilaian kemahiran generik dalam kursus pendidikan jasmani dan kesihatan (PJK): metode kajian. In *Prosiding Persidangan Antarabangsa Kelestarian Insan 2014 (INSAN2014)* (pp.225-236). Batu Pahat: Universiti Tun Hussein Onn Malaysia. Retrieved from <http://eprints.uthm.edu.my/5482/>
- Turner, J. (2001). Examining an Art Portfolio Assessment Using a Many-Facet Rasch Measurement Model. *Tesis Ed.D.* University of Northeastern.
- Uusikylä, K. (2012). *Luovuus kuuluu kaikille Creativity is for everyone*. Jyväskylä: PS- kustannus.
- Using Rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety. *JALT Journal*, 35(1), 5-28. Retrieved from <http://jalt-publications.org/files/pdf-article/jj35.2-art5.pdf>
- Vygotsky, L. (1978) *Mind in Society*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Wang, W. (2008) Cultural differences: why do Asians avoid extreme responses? *Survey Practices*, 1(3), 1-7. Assessment of differential item functioning. *Journal of Applied Measurement*, 9(4), 387-408.
- Wortham S.C., (2012). *Assessment in Early Childhood Education*. Ed. Ke -6. United States: Pearson.
- Wiggins., G. P.,(1993). *Assessing student performance*. San Francisco: Jossey-Base.
- Wright, B. D., & Panchapakesan. (1969). Procedure for sample free item analysis *Educational and Psychological Measurement*, 29, 23-48

Wright, B. D., & Masters. (1982). *Rating scale analysis Rasch measurement*. Chicago: MESA Press.

Wright, B. D., & Stone, (1979). *Best test design*. Chicago: MESA Press.

Wright, B. D., & Stone, (1979). *Best test design*. Chicago: MESA Press. Hambleton & Jones, 1993. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement : Issues and Practice*, 12(3), 38-47.

Wright, B. D., & Masters .(1982). *Rating scale analysis Rasch measurement*. Chicago: MESA Press.

University of Malaya

## Senarai Penerbitan

**Nor Mashitah, M. R.**, Mariani, M. N., Jain C., Mohamad Ilmee, M. Z., Hafiza S., & Rosmah A.G. (2014). Penilaian Perkembangan Holistik Kanak-kanak Berasaskan Pentaksiran Prestasi. *Jurnal Pendidikan Awal Kanak-Kanak*, Volume 3. Submitted: 10th September 2014. (Non-ISI/Non-SCOPUS)

**Nor Mashitah M.R.**, Mariani M.d, Jain Chee, Mohammad Ilmee Mohammad Zin, Hafiza Sulaiman, Rosmah A.G. (2015). Penggunaan Model Pengukuran Rasch Many-Facet (MFRM) Dalam Penilaian Kanak-kanak Berasaskan Prestasi. *Jurnal Pendidikan Awal Kanak-Kanak*. 4(9), 1-16 (Non-ISI/Non-SCOPUS)

Jain Chee, Mariani Md Nor, **Nor Mashitah Mohd Radzi**, Che Mah Yusof, Abdullah Kadir & Shahrul Nizam. (2015). The Preschool Curriculum Implementation Versus Problems. *Journal of Early Childhood Educations and Care* Vol 4, 2015. National Child Development Research Centre (NCDRC). Universiti Pendidikan Sultan Idris. Non-ISI (Non-ISI/Non-SCOPUS)

**Nor Mashitah M.R.**, Mariani M.N, Jain Chee & Che Mah Binti Yusuf. 2015. Developing a Measure of Authentic Assessment Standard for Children's Development and Learning Using Many Facet Rasch Model, *International Journal of Early Education and Care* Vol.4, Non-ISI/Non-SCOPUS Cited Publication (Non-ISI/Non-SCOPUS)

Mariani, M.N., Seri Banun, R., Norhashimi, Saad., Siti Fairos, S., **Nor Mashitah, M. R.**, Jain, C., Nur Zahirah, M. S., Che Mah, Y., Hafiza, S. 2016. School readiness among Indigenous children. *Taylor & Francis Group*, 15(30), 1-11. (SCOPUS-Indexed)

**Nor Mashitah M.R.** and Mariani M.N.,(2016). Measurement of Early Science and Mathematics Standards Instruments: Performance Assessment and Psychometric setting using ZPD Concept. *Indian Journal of Science and Technology*, 9(47),233-239 (SCOPUS-Indexed)

Mariani, M.N., Siti Fairos, S., Norhashimi, Saad., Seri Banun, R., **Nor Mashitah, M. R.**, Jain, C., Nur Zahirah, M. S., Che Mah, Y., Hafiza, S. 2016. Developing a systematic assessment system for the Linus programme. *Taylor & Francis Group*, 15(30), 25-36 (SCOPUS-Indexed)

**Nor Mashitah M.R.**, Abdul Halim Masnan, Md. Nasir Masran, Mariani M.N. (2016). Evaluation of an Early Science and Mathematics Standard Instrument Task Development, Scoring Set, and Psychometric Issues. *Advanced Science Letter*. (SCOPUS-Indexed)

Mohd Nazri Abdul Rahman, Mariani M.N., Noraziah Ahmad Nadzim, **Nor Mashitah M.R** and Nurnadia Moktar. (2016). Application of Fuzzy Delphi Approach in Designing Homeschooling Education for Early Childhood Islamic Education. *International Journal Of Academic Research In Business And Social Science (Era)*. 6(12). 837-846 (SCOPUS-Indexed)

Md. Nasir Masran, **Nor Mashitah M.R.**, Mariani M.N, (2016). Assessment of holistic child development based on Performance Assessment Using Many-Facet Rasch Model (MFRM). *Advanced Science Letter*. (SCOPUS-Indexed)

Md. Nasir Masran, **Nor Mashitah M.R.**, Mariani M.N. (2016). Validating Measure of Authentic Assessment Standard for Children's Development and Learning Using Many-Facet Rasch Model. *Advanced Science Letters* (SCOPUS-Indexed)

Jain Chee, Mariani M.N, Abdul Jalil Othman, **Nor Mashitah M.R.** (2017). Exploring the Issue of Content Knowledge, pedagogical and Technological Among Preschool Teachers. *International Journal of Advanced and Applied Sciences*, 4(3), 130-136. (SCOPUS-Indexed)

Jain Chee, Mariani M,N, Abdul Jalil Othman, **Nor Mashitah M,R.** (2017). Understanding of Content Knowledge Pedagogical Knowledge among Preschool Teacher and Application Developmentally Appropriate Practices in Teaching. *International Journal of Advanced and Applied Sciences*, 4(3), 148-153. (SCOPUS-Indexed)

**Nor Mashitah, M.R.** , Mariani M.N., Shahrir Jamaluddin & Mohd Nazri Abdul Rahman (2017). Using G-Theory in the Development of Performance Assessment of the Socio-Emotional Domain of Children. *International Journal of Advanced and Applied Sciences*, 4 (3), 141-147. (SCOPUS-Indexed)

Mohd Effendi @ Ewan Mohd Matore, Ahmad Zamri Khairani, Mariani M.N, **Nor Mashitah Mohd Radzi.** (2017). Evaluating The Content Validity of IKBAR using Content Validity Ratio Among Professional Experts. *International Electronic Journal of Mathematics Education*, 15(30), 5-13 (SCOPUS-Indexed)

Mohd Effendi @ Ewan Mohd Matore, Ahmad Zamri Khairani, Mariani M.N, **Nor Mashitah Mohd Radzi.** (2017). Psychometric Properties on ISIS in Measuring Spiritual Quotient among Polytechnic student using the Rasch Measurement Model. *The International of Environmental and Science Education*, 15(30), 25-36. (SCOPUS-Indexed)

**Nor Mashitah M.R.,** Mariani M. N, Shahrir Jamaluddin. (2017). Dependability Score For Developmental Of Creativity Domain Through Performance-Based Assessment Using Generalizability Theory. *The International Journal Of Environmental And Science Education*, 15(30), 35-44. (SCOPUS-Indexed)

**Nor Mashitah M.R.,** Mariani M. N, Mohd Effendi@Ewan Mohd Matore, (2017). Dependability Score For Developmental Of Language Domain Through Performance-Based Assessment Using Generalizability Theory. *The International Journal Of Environmental And Science Education*. 15(30), 44-54. (SCOPUS-Indexed)

**Nor Mashitah M.R,** and Mariani M.N. Using G-Theory in the Developmental of Performance Assessments of the Physical Domain of Children. *International Journal of Advanced and Applied Sciences*, 4(3), 154-159. (SCOPUS-Indexed)

**Nor Mashitah, M.R.,** Mariani M.N, , Mohd Effendi@ Ewan Mohd Matore. (2017). Using G-Theory in the Developmental of Performance Assessment of the Cognitive. *International Electronic Journal of Mathematics Education*, 15(30), 14-24. (SCOPUS-Indexed)