

PENGGUNAAN MODEL PENGUKURAN RASCH *MANY-FACET* (MFRM) DALAM PENILAIAN PERKEMBANGAN KANAK-KANAK BERASASKAN PRESTASI

Nor Mashitah M.R¹, Mariani M. N², Jain Chee³, Mohamad Ilmee Mohamad Zin⁴,
Hafiza Sulaiman⁵ & Rosmah A.G⁶
Universiti Malaya

ABSTRAK

Objektif kajian ini adalah untuk menilai perkembangan kanak-kanak berasas prestasi menggunakan Model Pengukuran Rasch *Many-Facet*. Data dikumpul daripada 60 orang kanak-kanak berumur 2 dan 3 tahun menggunakan 2 *rater*. Instrumen NoaMa digunakan untuk menilai prestasi kanak-kanak dalam enam domain. Model Pengukuran Rasch *Many-Facet* (MFRM) digunakan untuk menganalisis data untuk melihat tiga aspek iaitu *facet* kebolehan kanak-kanak, kesan *rater* dan kesukaran domain. Dapatan menunjukkan indeks yang tepat mengukur *facet* kanak-kanak berkebolehan tinggi dan berkebolehan rendah pada nilai ukuran 0.96 hingga -0.57, *facet* ketegasan *Rater* pada nilai ukuran 0.07 hingga -0.07, dan *facet* tahap kesukaran domain adalah nilai ukuran 0.96 (domain rohani) hingga -0.25 (domain fizikal). Nilai kebolehpercayaan dan kesahan konstruk yang tinggi mengikut MFRM menunjukkan model pengukuran ini dapat mengukur ketepatan skor setiap *facet*.

Kata kunci: perkembangan kanak-kanak, penilaian prestasi, *rater*, model pengukuran rasch *many-facet* (MFRM)

ABSTRACT

The objective of this study was to evaluate the development of children using performance-based Many-Facet Rasch Measurement model (MFRM). Data was collected from 60 children aged 2 and 3 years old with 2 raters. NOaMA instrument used to assess children's achievement in six domains. MFRM was employed to analyze the data looking at three aspects facets, namely the abilities of children, the effects of a rater and difficulties domain facets. The results showed that the index accurately measures high and low ability children facet on the size of 0.96 to -0.57, the rater severity facet on the size of 0.07 to -0.07, and difficulty domain facet on the size of 0.96 (spiritual domain) to -0.25 (physical domain). The high reliability and construct validity by MFRM showed it can measure the accuracy of the scores of each facet successfully .

KeyWords: children development, performance based, rater, Many-facet Rasch Measurement (MFRM)

PENGENALAN

Kebelakangan ini, berlaku peningkatan penggunaan alat penilaian untuk melihat kebolehan dan prestasi kanak-kanak seawal usia mereka. Peningkatan ini berlaku disebabkan pendekatan pendidikan masa kini menekankan perkembangan kanak-kanak secara menyeluruh atau holistik. Didikan lengkap ini mengambil kira keseluruhan dimensi perkembangan iaitu kognitif, bahasa, fizikal, sosioemosi dan rohani (Morrison, G.S, 2011).

Tahap kebolehan dan prestasi kanak-kanak diperolehi melalui peranan guru dengan menjalankan penilaian bagi setiap perkembangan. Ia bertujuan untuk menentukan tahap perkembangan kanak-kanak dan mengenal pasti kemahiran sedia ada atau yang belum dan perlu dikuasai. Walau bagaimanapun, tujuan penilaian tidak lagi mengambil kira keperluan dan minat kanak-kanak apabila mereka melangkah ke alam persekolahan (Wortham, 2012). Oleh itu, prinsip penilaian untuk peringkat kanak-kanak bukan sahaja berkaitan dengan penilaian semata-mata, tetapi memberi implikasi kepada bagaimana data kanak-kanak itu dianalisis dan menghasilkan hasil dapatan yang sah dan boleh dipercayai. Pada peringkat awal kanak-kanak, penilaian terhadap perkembangan mereka adalah fokus utama.

Penilaian yang berkesan merupakan satu tanggungjawab yang penting. Alat penilaian yang sesuai berupaya mengesan tahap kebolehan dan prestasi kanak-kanak. Pencapaian yang baik dan kanak-kanak yang berkebolehan memberi kesan positif kepada kualiti sesebuah taska dan perkembangan kanak-kanak khususnya. Perkara ini turut disokong oleh Sylva, Siraj-Blatchford, Taggart, Sammons, Melhuish, Elliot, & Totsika (2006), menyatakan kanak-kanak yang menerima pembelajaran yang berkualiti tinggi lebih berjaya pada peringkat sekolah rendah. Disebabkan kanak-kanak menunjukkan kebolehan yang tinggi dalam aspek sosioemosi, bahasa dan intelektual semasa peringkat awal kanak-kanak. Selain kepentingan penilaian diambil kira, prestasi kanak-kanak perlu dianalisis menggunakan model pengukuran yang sesuai dan berupaya memberi hasil analisis yang tepat dan adil untuk kanak-kanak secara individu dan tidak dibandingkan antara kanak-kanak.

Selari dengan pelbagai aspek dan masalah tersebut, Kajian ini menekankan penganalisan data menggunakan Teori Pengujian Moden iaitu menggunakan Model Pengukuran Rasch *Many-facet* (MFRM). Setiap aspek dikenali sebagai *facet* dan hasil analisis menunjukkan model ini mengukur setiap kanak-kanak secara individu. MFRM akan mengukur aspek kebolehan kanak-kanak, dan faktor *rater* (penilai), dan kesukaran domain secara kuantitatif. Nilai kesahan dan kebolehpercayaan penilaian prestasi yang dinyatakan merumuskan kebolehan sampel kanak-kanak yang dicatatkan dengan tepat menggunakan MFRM (Engelhard dan Stone, 1998).

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

Berdasarkan kajian ini, metodologi MFRM akan digunakan untuk menilai prestasi kanak-kanak melalui nilai skala rating *rater* yang mengenal pasti kebolehan kanak-kanak,

ketegasan *rater* dan kesukaran domain perkembangan sekaligus menghasilkan dapatan data yang dianalisis yang sah dan boleh dipercayai. Metodologi ini dapat dimanfaatkan oleh penyelidik-penyelidik dalam bidang pendidikan awal kanak-kanak dalam merancang, melaksanakan dan menilai perkembangan kanak-kanak yang mempertimbang untuk menggunakan *rater*. Hasil kajian ini dapat menunjukkan prestasi kanak-kanak dinilai secara kuantitatif, dipercayai, sah dan lebih objektif.

Kerangka Teori

Kajian ini berlandaskan Teori Perkembangan Piaget dan Teori Vygotsky. Kajian ini menghubungkan pembelajaran dan perkembangan kanak melalui penekanan kepada konsep kanak-kanak adalah individu yang dinamik dan berbeza tahap kebolehan.

Pembelajaran kanak-kanak melibatkan interaksi dinamik dengan semua aspek perkembangannya iaitu fizikal, sosioemosi, bahasa, kognitif 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.

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, penilaian perlu dilakukan semasa proses pembelajaran dimana kanak-kanak berinteraksi dengan guru dan rakan-rakan melakukan sesuatu aktiviti. Penilaian domain-domain perkembangan dapat dicatat dengan lebih maksimum.

Penilaian

Penilaian terhadap kanak-kanak dilakukan untuk mengenal pasti banyak perkara terutama untuk melihat tahap penerimaan pembelajaran dan proses perkembangannya. Penilaian

terhadap kanak-kanak dilakukan melalui pemerhatian. Kaedah pemerhatian berkait terus dengan pembelajaran dan perkembangan kanak-kanak. Fokus pemerhatian adalah kepada

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 menilai kebolehan kanak-kanak agar objektif yang ditetapkan dapat dicapai dan maklumat yang dihasratkan dapat diperoleh (Billman & Sherman, 1997; Harrington, 1997; Pelo, 2006).

Perkembangan Kanak-kanak

Definisi Konsep perkembangan 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).

Definisi Operasi dalam kajian ini adalah perkembangan kanak-kanak melalui beberapa peringkat untuk mencapai tahap maksimum. Penilaian perlu dilakukan untuk mengetahui kanak-kanak berada pada tahap mana kerana semakin sukar tugas yang diberikan semakin banyak bimbingan yang diperlukan. Melalui penilaian 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 mengambil kira definisi operasi ini. Perkembangan kanak-kanak yang dinilai secara holistik iaitu meliputi semua domain perkembangan iaitu domain kognitif, bahasa, fizikal, kreativiti, sosioemosi dan rohani. Definisi operasi perkembangan kanak-kanak dalam kajian ini merujuk kepada tiga perkara, iaitu:

a. Potensi Perkembangan didefinisikan sebagai mencapai tahap tertinggi suatu tugas berjaya dilaksanakan oleh kanak-kanak dengan bantuan dan bimbingan diberikan.

b. Perkembangan Sebenar didefinisikan sebagai mencapai tahap tertinggi suatu tugas berjaya dilaksanakan oleh kanak-kanak tanpa bantuan dan bimbingan.

c. 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. Sesetengah tugas mudah kepada kanak-kanak tapi tugas yang sama mencabar dan sukar bagi kanak-kanak yang lain walaupun dalam lingkungan umur yang berbeza.

Penilaian Prestasi

Definisi Konsep penilaian prestasi adalah penilaian yang memerlukan seseorang menunjukkan kemahiran dan pengetahuan yang dimilikinya sebagai maklum balas terhadap suatu tugas. Penilaian ini dilihat sebagai alternatif kepada pentaksiran aneka pilihan dan jawapan pendek yang hanya menumpukan kepada objektif pembelajaran (Russell & Airasian, 2012).

Definisi Operasi penilaian prestasi dalam kajian ini 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 penilaian prestasi yang terdapat dalam instrumen adalah meliputi keseluruhan enam domain perkembangan. Jadual 1 dan menunjukkan senarai domain perkembangan dan jumlah item.

Jadual 1
Senarai Domain Perkembangan dan Jumlah Bilangan

	Domain	Instrumen	Kod
(2-3 tahun)	1	Kognitif	K1-K26
	2	Bahasa	B1-B16
	3	Kreativiti	KR1-KR21
	4	Fizikal	F1-F39
	5	Sosioemosi	S1-S61
	6	Rohani	R1-R10
Jumlah bilangan			173

Soalan Kajian

- i. Apakah nilai kebolehppercayaan setiap *facet* kanak-kanak, *rater* dan domain perkembangan?
- ii. Apakah nilai kesahan setiap facet merujuk kepada nilai *Infit MNSQ* dan *Outfit MNSQ* bagi:

Kanak-kanak

Apakah ukuran yang dilaksanakan dapat mengasingkan kanak-kanak kepada keupayaan yang berbeza?

Rater

- i. Apakah nilai ukuran ketegasan *rater*?
- ii. Adakah terdapat perbezaan ketegasan?
- iii. Adakah terdapat konsistensi antara rater?

Facet domain perkembangan

- i. Di manakah kedudukan kesukaran setiap domain yang diukur?
- ii. Apakah ketepatan penentuan yang dibuat?

- iii. Apakah respons setiap domain sepadan (*fit*) mengikut jangkauan model pengukuran Rasch?

METODOLOGI

Rekabentuk 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 Merancang, Peringkat Membina Instrumen dan Peringkat Mentaksir. Proses mereka bentuk dilakukan selari dengan prosedur yang telah ditetapkan dalam Modul Penilaian Perkembangan Kanak-kanak yang dibina oleh Bahagian Pembangunan Kurikulum, Kementerian Pelajaran Malaysia pada tahun 2010. Prosedur ini diambil kira kerana modul penilaian ini adalah satu-satunya panduan yang dihasilkan mengikut budaya kanak-kanak Malaysia dan ia merupakan sistem yang telah dilaksanakan untuk menilai kanak-kanak di Malaysia.

Sampel kajian ini dipilih secara bertujuan daripada sembilan taska swasta. Instrumen NoaMA digunakan dan ditadbir kepada kanak-kanak berumur dua dan tiga tahun ($n:60$). Penilaian perkembangan kanak-kanak dalam kajian ini menggunakan penskoran secara skala likert (lima rating).

Instrumen NOaMA mematuhi konsep penilaian Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan telah adaptasi instrumen yang membahagikan kepada enam domain perkembangan seperti yang terdapat dalam Jadual 2, dan bermula tahun 2010, Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan telah digunakan di TASKA swasta berkaitan perkara berikut 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;

- vi. 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.

Jadual 2

Domain Perkembangan yang diadaptasi daripada Dasar Asuhan dan Didikan Awal Kanak-kanak Kebangsaan

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 dan 3 tahun)
Perkembangan Bahasa, Komunikasi dan Literasi Awal (Bahasa)	Domain Bahasa
Perkembangan Pra-Matematik dan Pemikiran Logik	Domain Kognitif
Perkembangan Sosio-emosi, Sahsiah dan Kerohanian (Rohani)	Domain Rohani
Perkembangan Fizikal dan Kemahiran Mengurus Diri (Fizikal)	Domain Fizikal
Perkembangan Estetika dan Daya Kreatif (Kreativiti)	Domain Kreativiti
Perkembangan Sosio-emosi, Sahsiah dan Kerohanian (Sosioemosi)	Domain Sosioemosi

Peringkat Merancang

Pada peringkat perancangan, kumpulan sasaran kanak-kanak terdiri kanak-kanak berumur 2 dan 3 tahun di Taman Asuhan Kanak-kanak (Taska) swasta dengan mengambil kira komposisi jantina dan lokasi Taska. 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 ia juga berupaya menentukan ranking mereka dengan tepat. Dua orang *rater* juga dilantik sebagai alat kajian yang melakukan pemerhatian sampel kajian. *Rater* yang sama juga melakukan pemerhatian di lokasi Taska yang berbeza.

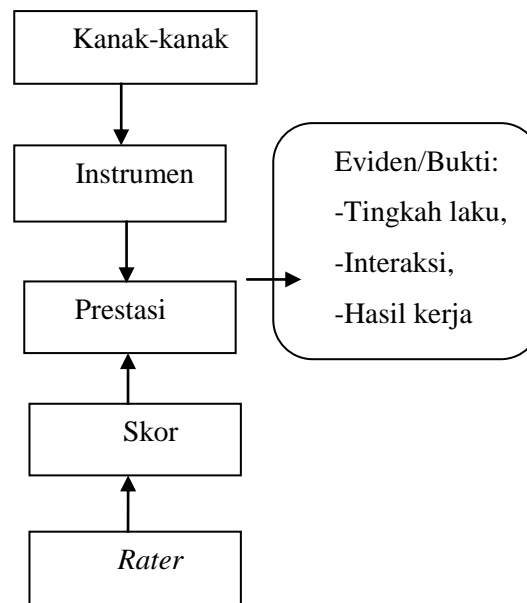
Peringkat Membina Instrumen

Pada peringkat pembinaan instrumen, instrumen telah direkabentuk semula dari segi susun atur kandungan instrumen, menggunakan skor skala rating dan menyusun atur domain-domain instrumen. Kajian rintis dijalankan untuk mendapatkan nilai kesahan dan kebolehpercayaan dan melaporkannya.

Pada peringkat penilaian, kaedah pemerhatian digunakan untuk mendapatkan eviden daripada prestasi yang ditunjukkan oleh kanak-kanak. Pemerhatian dilakukan oleh *rater* dan menggunakan borang untuk menentukan skala kebolehan kanak-kanak daripada 1 hingga 5.

Peringkat Mentaksir

Kajian ini menekankan penilaian prestasi kepada semua domain perkembangan dalam suasana pembelajaran yang menyeronokkan (*fun learning*) 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. Pendekatan ini boleh dilakukan tanpa kanak-kanak tahu bahawa dia diperhatikan. Kajian ini menggunakan peranan *Rater* untuk membuat pemerhatian terhadap kanak-kanak seperti Rajah 1 di bawah menunjukkan proses penilaian instrumen prestasi menggunakan *rater*.



Rajah 1 : Proses Penilaian guna *Rater*

Perkara yang diperhatikan ialah eviden atau bukti. Bukti prestasi yang boleh dipamerkan dalam bentuk tingkah laku, interaksi, hasil kerja. Bukti yang mahu dilihat oleh *rater* ialah yang menunjukkan perkembangan, pertumbuhan dan pencapaian kanak-kanak dari segi tingkah laku, interaksi dan hasil kerja kanak-kanak. *Rater* akan diberi latihan agar tiada jarak yang ketara antara mereka dan untuk meningkatkan konsistensi antara keduanya semasa penilaian.

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 rater melalui pemerhatian secara langsung dan hasil kerja. Rater 1 dan 2 akan mengisi borang Penilaian Domain Kanak-kanak (PDK) bagi setiap enam domain. Borang PDK mengandungi maklumat kriteria asas penilaian, item prestasi, skala 1 hingga 5. Kanak-kanak hanya akan menerima skor rating sahaja daripada *rater* dan dicatat dalam borang. Skala rating untuk PDK dan tindakan atau kebolehan kanak-kanak melakukan tugas berdasarkan item prestasi adalah:

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

Kerangka Statistik MFRM

MFRM digunakan sebagai kerangka statistik untuk menganalisis data daripada proses penilaian pembelajaran dan perkembangan kanak-kanak. Membuat keputusan berdasarkan data mentah yang dikumpul daripada skor prestasi tidak dapat menghuraikan kesukaran suatu domain, ketegasan setiap rater, dan tahap dimensi skala rating. Bagaimana kebolehan kanak-kanak dapat ditentukan dengan tepat, jika ketegasan *rater* terhadap domain yang sukar menentukan kanak-kanak A, dan kelembutan *rater* 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” MFRM 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 *rater*, menjadi lebih teratur kepada konstruk atau ciri terpendam (kebolehan kanak-kanak). Semakin tinggi skor, menunjukkan semakin besar jumlah ciri terpendam yang diukur. MFRM 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 MFRM menyediakan maklumat tentang menariknya dalam pengukuran, sama ada domain pembelajaran atau perkembangan terlalu sukar atau mudah, adakah *rater* tegas atau sebaliknya, kriteria yang dimiliki *rater* dalam keadaan yang konsisten, adakah kanak-kanak diberi skor tinggi kerana *rater* bersifat pemurah atau domain itu senang, kesemua soalan ini dapat dijawab semasa menggunakan model ini.

Model Rasch menyediakan pengkaji maklumat untuk ‘mengukur konstruk dengan selari daripada pemerhatian kualitatif’. Dengan menggunakan penilaian berasas prestasi, seperti hasil kerja, inilah model yang paling tepat diperlukan. Menggunakan *rater* 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 *rater* dengan kesukaran domain terhadap kanak-kanak, melihat interaksi ketegasan *rater* dengan kanak-kanak, dan skala rating yang digunakan oleh *rater*. Perbandingan kuantitatif oleh MFRM tidak bergantung ke atas siapa kanak-kanak yang ditaksir oleh mana-mana *rater* 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 rater dan setiap item.

Model Rasch adalah model statistik digunakan sebagai *confirmatory analysis*. Kriteria untuk berjaya 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 *rater* 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.

DAPATAN KAJIAN

Analisis pertama

Memfokuskan kepada nilai kebolehpercayaan. Nilai *Alpha Cronbach* kajian ini ialah bagi kanak-kanak (0.99), *rater* (0.97), dan domain perkembangan (1.00). Jadual 3 menunjukkan nilai kebolehpercayaan kanak-kanak, *rater* dan domain perkembangan bagi kajian ini adalah di antara 0.97 hingga 1.00 dengan indeks pemisahan di antara 6.02 hingga 23.30. Nilai ini adalah memenuhi kehendak model pengukuran Rasch. Nilai kebolehpercayaan responden ≥ 0.8 dan nilai indeks pemisahan ≥ 2 adalah indeks yang diterima baik.

Jadual 3
Nilai kebolehpercayaan dan indeks pemisahan bagi 3 *facet* Instrumen NoaMA

<i>Facet</i>	Nilai Kebolehpercayaan	Indeks Pemisahan
Kanak-kanak	0.99	9.81
Rater	0.97	6.02
Domain Perkembangan	1.00	23.30

Analisis kedua

Melihat nilai kesahan setiap *facet* 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 0.4 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 0.6 hingga 1.4 adalah nilai julat yang 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 MFRM. Nilai kesahan kajian ini akan dinyatakan dalam huraian hasil dapatan setiap *facet*.

Facet kanak-kanak

i. Apakah ukuran yang dilaksanakan dapat mengasingkan kanak-kanak kepada keupayaan yang berbeza?

Jadual 4 menunjukkan laporan ukuran kanak-kanak iaitu nilai ukuran kanak-kanak, nilai RMSE, ralat piawai (SE), nilai indeks pemisahan 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 *facet* kanak-kanak adalah 0.07. Nilai indeks pemisahan adalah 9.81 menunjukkan kanak-kanak ini dapat dipisah atau diasingkan

kepada keupayaan yang berbeza. Ini disahkan lagi dengan nilai khikuasdua yang signifikan di mana bagi nilai $\chi^2 = 5759.5$, $p < 0.01$, $df = 59$. Nilai indeks pemisahan dan nilai khikuasa dua yang signifikan juga menunjukkan bahawa ukuran kanak-kanak yang diperolehi dapat memisahkan kanak-kanak kepada keupayaan yang berbeza.

Nilai kesahan merujuk statistik *Infit MNSQ* dan *Outfit MNSQ* yang disemak untuk mengenal pasti data yang *fit* (sesuai) dengan model pengukuran ini. Hasil Dapatan menunjukkan bacaan nilai *Infit MNSQ* (0.90 – 1.28) dan *Outfit MNSQ* (0.85 – 1.30) adalah dalam julat yang diterima dan menunjukkan data *facet* kanak-kanak memenuhi jangkaan model pengukuran ini seterusnya mempunyai nilai kesahan.

Jadual 4
Kebolehan kanak-kanak

Kanak-kanak	Nilai ukuran	Ralat Piawai (SE)	<i>Infit MNSQ</i>	<i>Outfit MNSQ</i>
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-kuasdua: 5759.5 ; df : 59; Signifikan (kebarangkalian) : 0.00
Nilai RMSE: 0.07; Nilai Pemisahan Indeks: 9.81

Facet Rater

i. Ukuran Ketegasan

Jadual 5 menunjukkan statistik pengukuran *facet rater* terdiri daripada nilai ukuran ketegasan, ralat piawai dan nilai *infit* dan *outfit* MNSQ. Ketegasan rater berada pada julat -0.07 logit (SE= 0.01) bagi Azwani iaitu *rater* yang pemurah (*lenient*) sehingga 0.07 logit (SE= 0.01) bagi rater Azila iaitu *rater* yang tegas. Nilai ralat piawai yang dinyatakan adalah sebagai indikator ketepatan bagi setiap ukuran ketegasan.

ii. Perbezaan ketegasan

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

iii. Konsistensi rater

Nilai *infit* dan *outfit* MNSQ bagi setiap *rater* dalam menilai perkembangan kanak-kanak adalah menggambarkan konsistensi *rater* dalam membuat penilaian. Nilai *Infit* dan *Outfit* MNSQ bagi setiap *rater* adalah berada dalam julat yang boleh diterima iaitu 0.6 dan 1.4. Iaitu Azila (*Infit* MNSQ: 0.99, *Outfit* MNSQ: 0.98) dan Azwani (*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 5
Pengukuran *Facet Rater*

Rater	Ukuran Ketegasan	Ralat (SE)	Piawai	<i>Infit MNSQ</i>	<i>Outfit MNSQ</i>
Azila	0.07	0.01		0.99	0.98
Azwani	-0.07	0.01		1.04	1.03

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

Facet domain perkembangan

i. Di manakah kedudukan kesukaran setiap domain yang diukur?

Berdasarkan Jadual 6, 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. Rajah 2 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.

ii. Apakah ketepatan penentuan yang dibuat?

Ketepatan pengukuran dilihat dari nilai ralat piawai (SE) dalam Jadual 6. Nilai ralat piawai bagi semua domain berada di antara 0.01 hingga 0.03.

iii. Apakah respons setiap domain sepadan (*fit*) mengikut jangkaan model pengukuran Rasch?

Jadual 6 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). Kesemua domain didapati mempunyai nilai yang sepadan dan memenuhi model pengukuran Rasch seterusnya mempunyai nilai kesahan. Jadua

Pengukuran Domain Perkembangan

Domain	Nilai ukuran	Logit SE	<i>Infit MNSQ</i>	<i>Outfit MNSQ</i>
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

Ukuran	Kanak-kanak	Rater	Domain	Item	Skala
3	+	+		+	(5)
	**				
2	+	+		+	---
	**				
	*				
	**				

	**		*		

	***			4	
	**				
1	+	+	Rohani	+	*

	***			*	
	*				
	****			***	---
	****			****	
	***			**	
	**			**	
	**	Bahasa	Sosioemosi	*****	
	**	Azila		***	3
* 0	*	*		*****	* *
	**	Azwani		*****	
	*	Kreativiti		*****	
	*	Kognitif		*****	

				***	---
	*			****	
		Fizikal		*	
-1	+	+		+	(1)
Ukuran	* = 1	Rater	Domain	* = 1	Skala

Rajah 2 : Kedudukan Domain Mengikut Kesukaran

PERBINCANGAN

Hasil dapatan analisis pertama, Nilai kebolehpercayaan bagi facet kanak-kanak, *rater* 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 pemisahan (*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 pemisahan yang lebih tinggi bagi unsur-unsur dalam faset tersebut (Wright & Masters, 1982). Dapatan menunjukkan semua *facet* mempunyai nilai kebolehpercayaan yang tinggi.

Analisis kedua, bagi *facet* kanak-kanak, Jadual 4 menunjukkan nilai indeks pemisahan adalah 9.81 menunjukkan kanak-kanak ini dapat dipisah atau diasingkan kepada keupayaan yang berbeza. Bagi *facet rater*, dalam Jadual 5 penyelidik menjawab dengan melihat hasil analisis FACETS berdasarkan model pengukuran Rasch terhadap ukuran ketegasan *rater* serta nilai khi-kuasadua bagi menguji sama ada terdapat perbezaan ketegasan yang signifikan antara *rater*. Nilai statistik kesepadanan pula digunakan untuk menentukan kesesuaian data *rater* membuat penilaian (Linacre, 2014). *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*. Seterusnya, Jadual 6 adalah *facet* domain perkembangan. Dapatan menunjukkan terdapatnya perbezaan yang signifikan dalam kesukaran bagi antara domain-domain. Rajah 2 turut menunjukkan kedudukan setiap domain yang menjelaskan tahap kesukaran domain hingga tahap yang paling mudah. Selain itu, ketepatan pengukuran dilihat daripada nilai ralat piawai (SE) berdasarkan Jadual 6 tersebut.

Seterusnya, nilai kesahan merujuk kepada nilai min kuasadua (MNSQ) dalam jadual 4,5 dan 6 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. *Rater* yang mempunyai nilai kurang daripada 0.6 ataupun melebihi daripada 1.4 adalah tidak konsisten dalam membuat penilaian. Dapatan telah menunjukkan *rater* 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.

KESIMPULAN

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

RUJUKAN

- Billman, J., & Sherman, J.A., (1997). *Observation and participation in early childhood settings*. Boston: Allyn & Bacon.
- Engelhard G., Stone G. (1998) Evaluating the quality of ratings Obtained from standard-setting judges. *Educational & Psychological Measurement*, 58(2),179-196.
- Harrington, H.L., Meisels, S.J., McMahon, P., Dichtelmiller, M.L., & Jablon, J.R. (1997). *Observing, documenting, and assessing learning: The work sampling system handbook for teacher educators*. Ann Arbor, MI: Rebus.
- Kondo-Brown, K. (2002). A FACET analysis of rater bias in measuring Japanese second language writing prestasice. *Language Testing*, 19, 3-31
- Linacre, J. M. (2003). The hierarchical rater model from a Rasch perspective. *Rasch Measurement Transactions*, 17, 928.
- Linacre, J.M. (1994). *Many-facet Rasch Measurement*. Chicago: University of Chicago, Measurement, Evaluation, and Statistical Analysis (MESA) Press.
- Linacre, J.M. (2014). *A User's Guide to FACETS Rasch-Model Computer Programs*. Program Manual 3.71.4.
- McNamara T.F (1996). *Measuring Second Language Prestasice*. London: Longman.
- Morrison G.S. (2011). *Early Childhood Education Today*. Ed. ke-12. Prentice Hall: New Jersey.

- MyFord, C.M. & Mislevy, R.J. (1995). *Monitoring and improving a portfolio assessment system*. Princeton, New Jersey: Educational Testing Service.
- Myford, C.M & Wolfe, E.W. (2002). When rater disagree, then what? Examining third-rating discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement*, 3, 300-324.
- Pelo, A. (2006). Growing a culture of inquiry: Observation as professional development. *Exchange*, 172, 50-53.
- Russell M.K. & Airasian P.W. (2012). *Classroom Assessment: Concepts and Applications*. Edisi ke-7. McGrawHill: New York.
- Sylva K., Siraj-Blatchford I., Taggart B., Sammons P., Melhuish E., Elliot K., & Totsika V., (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, 21, 76-92.