

Pembangunan Instrumen Kemahiran Generik Pelajar Berasaskan Penilaian Pensyarah dengan Menggunakan Model Pengukuran Rasch Pelbagai Faset (The Developing Students Generic Skills Instrument Through Lecture Assessment Based on Many-Facets Rasch Measurement Model)

SITI RAHAYAH ARIFFIN, JAMIL AHMAD & NUR'ASHIQIN NAJMUDDIN

ABSTRAK

Kemahiran generik adalah sesuatu yang subjektif sifatnya. Oleh itu, pengukurannya boleh dibuat melalui rangkaian pembuktian seperti penilaian sendiri, rakan sebaya, guru atau pensyarah dan penggunaan portfolio. Kajian ini akan menghasilkan satu instrumen dengan menggunakan salah satu daripada aspek pengukuran tersebut, iaitu penilaian pensyarah. Instrumen ini mengandungi sembilan konstruk kemahiran, iaitu kemahiran kerja berpasukan, berkomunikasi, kepimpinan, pemikiran kritis dan penyelesaian masalah, pembelajaran sepanjang hayat dan pengurusan maklumat, keusahawanan, etika, moral dan profesional, sosial dan pengurusan. Kajian ini menggunakan Model Pengukuran Rasch Pelbagai-Faset (Many-Facets Rasch Measurement) yang melibatkan tiga faset, iaitu pelajar, item dan pensyarah penilai serta dianalisis dengan menggunakan perisian Facets 3.64.0. Penilai terdiri daripada 17 orang pensyarah dan subjek yang dinilai ialah 194 orang pelajar tahun 1 ambilan 2008/2009 di sebuah institusi pengajian tinggi. Instrumen yang telah ditadbir menunjukkan kebolehpercayaan bagi kesemua sembilan konstruk tersebut berada di antara 0.95-0.96. Item-item juga disemak dari segi keserasian data dengan Model Pengukuran Rasch dengan melihat nilai outfit mean square dan infit mean square. Peratus persetujuan antara penilai adalah di antara 43.4% – 49.5% dan menunjukkan nilai yang melebihi daripada jangkaan Model Pengukuran Rasch. Konsistensi penilai dalam membuat penilaian juga dilihat melalui analisis fit. Berdasarkan dapatan awal yang diperolehi mendapati instrumen yang dibina sesuai untuk digunakan bagi mengukur kemahiran generik pelajar berasaskan perspektif pensyarah.

Kata kunci: Kemahiran generik, pengukuran Rasch pelbagai faset, penilaian, kebolehpercayaan, kesahan

ABSTRACT

Generic skills are subjective in nature, therefore they can be measured by means of self-assessment, through the peers, teachers or lecturers and by using portfolio. This study aims at developing an instrument measuring the generic skills of university students using an approach from the evaluation of the lecturers. The instrument consists of nine construct of the generic skills, namely team-work, communication, leadership, critical thinking and problem solving, life-long learning and information management, entrepreneurships, ethics, moral and professional, social responsibilities and management. This study employs the Many-Facets Rasch Measurement Model, whereby it involves three facets namely the students, the items and the lecturers. The data analysis is done using Facets 3.64.0. The raters are 17 lecturers and the subjects rated are 194 first year students of 2008/2009 in-take in a university. The reliability scores of the nine constructs in the instrument are between 0.95-0.96. All items are also examined to determine whether the data fit the Rasch Measurement Model. This is done by means of looking at the outfit mean square values and the infit mean square values. The percentage of agreement amongst the raters is between 43.4% – 49.5%. This percentage range shows the values exceeding the expectation of the Rasch Model. The consistency among raters when making the rating is also examined using the fit analysis. Based on the findings from the analysis, it can be concluded that the instrument constructed is suitable for measuring the generic skills of university students through the perspectives of the lecturers.

Keywords: Generic skills, many-facets Rasch measurement, raters, reliability, validity

PENGENALAN

Satu instrumen yang sah dan boleh dipercayai perlu dihasilkan bagi menentukan tahap kemahiran generik pelajar Institut Penagajian Tinggi (IPT) kerana kemahiran generik adalah antara kemahiran yang sangat diperlukan dalam

pekerjaan (Siti Rahayah 2003). Selain daripada menimba ilmu pengetahuan dalam bidang yang diceburi, pelajar IPT juga perlu memiliki kemahiran generik untuk melengkapkan ciri modal insan yang dikehendaki oleh Misi Nasional dan Rancangan Malaysia Kesembilan. Misi nasional dan Rancangan Malaysia Kesembilan antara lain memberi

penekanan terhadap usaha meningkatkan keupayaan pengetahuan, kreativiti dan inovasi serta memupuk minda kelas pertama. Oleh itu, satu instrumen yang sah dan boleh dipercayai adalah perlu untuk mengukur kemahiran generik pelajar agar kemahiran tersebut dapat dikenal pasti dan ditingkatkan sebelum pelajar melangkah ke alam pekerjaan (Siti Rahayah 2008).

Terdapat pelbagai kaedah bagi pelajar untuk mengetahui kemahiran generik yang ada pada pelajar dengan bantuan para pensyarah atau guru. Antara kaedah penilaian yang boleh digunakan dalam menilai kemahiran generik adalah melalui ujian piawai pelbagai pilihan, penilaian sendiri, penilaian oleh rakan sebaya, portfolio dan juga melalui perspektif atau penilaian oleh pensyarah dan penilai (Grummon 1997). Penilaian oleh pensyarah juga kadangkala dikenali sebagai penilaian berasaskan kecekapan atau penilaian autentik. Penilaian begini juga biasa digunakan, terutamanya di tempat kerja dengan tujuan untuk menilai kemahiran. Bagi kajian ini, pelajar dinilai oleh pensyarah dengan menggunakan kriteria pemarkahan apabila pelajar mempamerkan kemahiran daripada aktiviti yang mereka lakukan di dalam kelas (McCurry & Bryce 1997; 2000).

Menurut Popham (1995), penilaian berasaskan kecekapan telah mula mendapat perhatian kerana pertama, penilaian melalui ujian piawai pelbagai pilihan adalah sangat terhad dan pelajar tidak dapat mempamerkan kemahiran sebenar yang mereka miliki. Kedua, terdapat kesedaran bahawa kebanyakan pelajar yang meninggalkan sekolah tidak dapat mengaplikasikan pengetahuan isi kandungan di luar bilik darjah, kerana kebanyakan ujian yang ada tidak meminta pelajar mendemonstrasikan kemahiran yang mereka ada, tetapi sebaliknya hanya menterjemahkannya dalam bentuk ujian pelbagai pilihan. Ketiga, dalam ujian yang dijalankan, guru atau pensyarah biasanya menitikberatkan maklumat dan strategi yang diperlukan bagi menjawab ujian tersebut. Keempat, apabila ujian hanya terdiri dari jawapan pelbagai pilihan, guru mungkin tidak membantu pelajar bagaimana untuk mengaplikasikan pengetahuan tersebut. Sekiranya penilaian memerlukan kecekapan atau aplikasi maklumat, guru atau pensyarah akan cuba melibatkan aktiviti yang dapat mengaplikasikan maklumat tersebut di dalam aktiviti kelas (Shavelson & Baxter 1992).

Instrumen berbentuk penilaian oleh penilai ini pernah dibina oleh McCurry dan Bryce (1997; 2000) di mana mereka telah dapat membentuk satu panel penilai dan memberikan mereka kursus yang bersesuaian dalam kemahiran generik untuk membolehkan mereka memberi pertimbangan yang konsisten terhadap kebolehan pelajar dalam kemahiran generik. Kursus dan pemerhatian yang dilakukan terhadap pelajar, sama ada berasaskan kelas atau aktiviti kokurikulum, membolehkan penilai memberikan pertimbangan yang konsisten untuk membezakan tahap kebolehan pelajar. Walau bagaimanapun, instrumen ini tidak diuji akan kesahan daripada aspek penilai walaupun menggunakan penilai dalam membuat penilaian terhadap kemahiran generik pelajar. Terdapat juga instrumen kemahiran generik lain seperti yang dibina oleh Iramaneerat (2008) untuk mengukur

kemahiran komunikasi, Matsuno (2006) bagi mengukur kemahiran penulisan, Banno (2008) bagi kemahiran perbualan di mana mereka hanya mengukur satu konstruk kemahiran generik sahaja.

TUJUAN KAJIAN

Kajian ini adalah bertujuan menghasilkan satu instrumen kemahiran generik penilaian pensyarah yang mempunyai kesahan dan kebolehpercayaan yang memenuhi Model Pengukuran Rasch Pelbagai Faset – PRPF (Linacre 1994).

METODOLOGI

Pembinaan instrumen dilaksanakan mengikut prosedur asas yang melibatkan 10 langkah yang dibahagikan kepada tiga peringkat, iaitu peringkat reka bentuk, pembinaan dan pengesahan. Pada peringkat reka bentuk, penyelidik membahagikan kepada empat langkah, iaitu tujuan pembinaan instrumen, menentukan kandungan instrumen, memberi definisi dan membina jadual spesifikasi instrumen. Pada peringkat pembinaan, penyelidik membahagikannya kepada tiga langkah, iaitu penulisan item, penghakiman panel pakar dan semakan item. Bagi peringkat pengesahan pula, terdapat tiga langkah, iaitu kajian rintis, analisis kebolehpercayaan dan kesahan dan penambahbaikan item (Cohen & Swerdlik 2005; Downing 2006; Gregory 2007; Wright & Stone 1979).

Pengkaji telah meneliti empat model kemahiran generik daripada buah empat negara, iaitu Amerika Syarikat, Australia, Britain dan Malaysia. Di Amerika digunakan *Secretary's Commission of Achieving Necessary Skills* (SCANS) yang menjalankan kajian *What work requires of schools* untuk Jabatan Buruh Amerika Syarikat (SCANS 1991) dan Model Carnavale (Carnavale, Gainer & Meltzer 1990). Dari Australia digunakan Model Mayer (Mayer Committee 1992). Model Kearns (Kearns 2001) dari Britain yang terdiri daripada *key skills, core skills* (1990) serta Model Kemahiran Generik oleh Bennet, Dunne dan Carre (1999) dan juga berdasarkan Kerangka Kelayakan Malaysia (KKM) (Sharifah Hapsah 2005) telah menghasilkan sembilan konstruk kemahiran generik, iaitu 1) kemahiran komunikasi, 2) kepimpinan, 3) kerja berpasukan, 4) pembelajaran sepanjang hayat dan pengurusan maklumat, 5) pemikiran kritis dan penyelesaian masalah, 6) etika, 7) moral dan profesional, 8) pengurusan, keusahawanan dan 9) tanggungjawab sosial (Siti Rahayah & Nor Azaheen 2009). Setiap konstruk dibina berasaskan teori behaviorisme, kognitivisme dan konstruktivisme seperti yang disarankan oleh Ertmer dan Newby (1993) di mana menurut mereka behaviorisme menghasilkan pembelajaran dengan perubahan sama ada dalam bentuk frekuensi atau kecekapan yang boleh diperhatikan.

Untuk menentukan kesahan dan kebolehpercayaan instrumen, satu kajian telah dijalankan kepada 194 pelajar.

Pelajar ini dinilai oleh 17 orang pensyarah penilai. Pemerhatian penilaian dijalankan selama tiga bulan. Setiap pelajar dinilai oleh tiga orang pensyarah penilai dengan menggunakan pelan penilaian yang telah ditetapkan bagi memenuhi Model PRPF. Pensyarah penilai kemudiannya mengisi instrumen berdasarkan pemerhatian telah dijalankan. Dapatan daripada kajian kemudiannya dianalisis dengan menggunakan perisian Facets 3.64.0 (Linacre 2008) untuk memastikan bahawa instrumen tersebut mempunyai kesahan dan kebolehpercayaan daripada aspek pelajar, item dan pensyarah penilai.

Model statistik yang digunakan dalam kajian ini adalah salah satu daripada Model Pengukuran Rasch, iaitu Model PRPF yang dibangunkan oleh Linacre (1994). Pembangunan Model PRPF ini telah mengembangkan potensi untuk menyelidik kesan daripada pelbagai sumber bagi perubahan yang ada dalam konteks penilaian prestasi. Hal ini membolehkan penilaian yang bersifat subjektif seperti penilaian prestasi ini menghasilkan satu penilaian yang objektif. Model ini adalah lanjutan daripada Model Pengukuran Rasch Satu Parameter yang telah dibangunkan untuk data dikotomis oleh George Rasch dalam tahun 1960-an (Wright & Masters 1999). Model ini menyediakan satu kerangka bagi menghasilkan pengukuran yang adil untuk keupayaan calon supaya tidak berubah dari segi statistik dalam kaitannya terhadap pelbagai faset yang terlibat dalam penilaian prestasi. PRPF menghasilkan satu kerangka untuk memeriksa kualiti psikometrik, iaitu kebolehpercayaan dan kesahan terhadap penilaian keupayaan yang diperolehi daripada penggunaan instrumen yang dihasilkan yang melibatkan pelbagai faset. Penggunaan PRPF ini telah mula berkembang luas, terutamanya dalam penilaian yang melibatkan pelbagai faset. Linacre (1994) menyatakan bahawa dengan PRPF, pembinaan sistem pengukuran bagi ujian yang bersifat subjektif adalah lebih praktikal dan berguna. Linacre (1994) juga menyatakan bahawa pembinaan-pembina ujian tidak perlu menghadkan soalan ujian hanya pada ujian yang berbentuk jawapan pelbagai pilihan sahaja (MCQ), tetapi juga menggunakan idea kreatif mereka bagi menghasilkan ujian yang melibatkan penilaian yang lebih

mendalam, autentik dan bersesuaian serta dapat mengekalkan kelebihan pengukuran yang objektif. Menurut Linacre (1994) lagi, sumbangan bagi setiap unsur bagi setiap faset terhadap sesuatu pemerhatian adalah didominasi oleh satu parameter dengan nilai yang bebas daripada parameter yang lain dalam satu rangka rujukan. Ini bermaksud, perbandingan bagi skor seorang pelajar boleh dilakukan tidak kira akan kesukaran item yang digunakan atau ketegasan penilai yang berbeza.

Pemilihan model pengukuran ini juga adalah kerana kesedaran penyelidik akan sifat penilai yang mempunyai ketegasan yang berbeza apabila menilai pelajar. Terdapat banyak kajian yang menggunakan penilaian prestasi bagi membetulkan ketegasan penilaian yang dibuat (Engelhard 1994, 1996, 1997; Engelhard & Myford 2003; Lunz et al. 1990; Lunz & O'Neil 1996; MacMillan 2000; Myford & Wolfe 2002). Kajian yang dijalankan oleh Lunz, Wright dan Linacre (1990) menunjukkan bahawa antara kelebihan model PRPF adalah model ini menyediakan satu kerangka rujukan bagi mengukur kesemua aspek dalam peperiksaan. Sebagai contoh, model PRPF telah digunakan untuk memeriksa kesan penggunaan penilai terhadap skor calon termasuklah kesan ketegasan dalam memberi skor terhadap calon (Kondo-Brown 2002; Weigle 1998; Engelhard 1994).

DAPATAN KAJIAN DAN PERBINCANGAN

Jadual 1 menunjukkan nilai kebolehpercayaan individu bagi kajian ini adalah di antara 0.94 dan 0.97 dengan indeks pengasingan di antara 3.97 hingga 5.51. Nilai kebolehpercayaan item pula berada di antara 0.82 hingga 0.99 dan indeks pengasingan di antara 2.12 hingga 0.99. Nilai ini adalah memenuhi kehendak Model Pengukuran Rasch. Linacre (2006) yang menyatakan kebolehpercayaan responden ≥ 0.8 dan indeks pengasingan responden ≥ 2.0 adalah indeks yang diterima baik (Siti Rahayah, Rodiah & Noriah 2010).

Kesahan dalam Model Pengukuran Rasch dilihat daripada nilai *fit* bagi setiap item. Item yang tidak *fit*

JADUAL 1. Nilai Kebolehpercayaan dan Indeks Pengasingan bagi Individu dan Item

KONSTRUK	INDIVIDU		ITEM	
	Kebolehpercayaan	Indeks Pengasingan	Kebolehpercayaan	Indeks Pengasingan
Komunikasi	0.95	4.23	0.95	4.35
Kepemimpinan	0.96	5.07	0.96	4.91
Kerja Berpasukan	0.96	5.03	0.97	5.40
Pembelajaran Sepanjang Hayat & Pengurusan Maklumat	0.97	5.51	0.94	3.98
Pemikiran Kritis & Penyelesaian Masalah	0.94	3.97	0.88	2.69
Etika & Moral Profesional	0.95	4.31	0.98	7.63
Pengurusan	0.96	4.81	0.96	4.68
Keusahawanan	0.96	4.86	0.82	2.12
Tanggungjawab Sosial	0.95	4.67	0.99	8.64

merupakan item yang perlu dimodifikasi atau dibuang. Item-item yang *fit* menunjukkan item yang boleh mengukur satu konstruk yang sama (Siti Rahayah et al. 2008h; Purya 2008). Menurut Linacre (2002), nilai *fit* yang boleh diterima dalam pengukuran adalah di antara 0.5 – 1.5. Nilai *point measure correlation* yang positif menunjukkan item bergerak selari (Bond & Fox 2001). Nilai *standard error* menunjukkan ketepatan dalam pengiraan (Siti Rahayah 2008d, 2008f; Linacre 2005; Wright 1995) dan nilai *measure* bagi item menunjukkan nilai kesukaran item-item bagi setiap konstruk. Bagi kajian ini, nilai-nilai yang diperolehi seperti dalam Jadual 2 menunjukkan item-item dalam setiap konstruk mempunyai kesahan yang memenuhi model PRPF.

Skala pemeringkatan pula diperiksa dengan menggunakan empat kriteria (Rodiah, Siti Rahayah & Noriah 2009; Linacre 1997). Pertama, setiap pemerhatian mestilah mempunyai sekurang-kurangnya 10 pemerhatian bagi setiap kategori. Kedua, *average category measure* menunjukkan peningkatan. Ketiga, nilai *outfit mean-square* mestilah kurang daripada 2.0. Keempat, *step difficulties* juga menunjukkan nilai yang meningkat dan nilainya mestilah meningkat sekurang-kurangnya 1.4 logit dan tidak melebihi 5.0 logit (Ling, Siti Rahayah & Sameah 2010; Bond & Fox 2001; Linacre 1997). Bagi kajian ini, kesemua konstruk menunjukkan skala pemeringkatan yang memenuhi model PRPF.

JADUAL 2. Nilai *Outfit* dan *Infit Mean Square*, *Standard Error*, *Point Measure Correlation* dan *Measure* bagi Sembilan Konstruk Kemahiran Generik

Konstruk	<i>Outfit MNSQ</i>	<i>Infit MNSQ</i>	<i>Standard Error</i>	<i>Point Measure Correlation</i>	<i>Measure</i>
Komunikasi	0.83 – 1.39	0.84 – 1.4	0.06 – 0.07	> 0.59	-0.43 – 0.56
Kepemimpinan	0.80 – 1.14	0.80 – 1.13	0.07	> 0.55	-0.63 – 0.36
Kerja Berpasukan	0.93 – 1.18	0.93 – 1.18	0.07	> 0.67	-0.52 – 0.63
Pembelajaran Sepanjang Hayat & Pengurusan Maklumat	0.96 – 1.09	0.96 – 1.06	0.07	> 0.68	-0.46 – 0.45
Pemikiran Kritis & Penyelesaian Masalah	0.86 – 1.41	0.86 – 1.43	0.06	> 0.56	-0.28 – 0.30
Etika & Moral Profesional	0.80 – 1.45	0.86 – 1.40	0.06 – 0.07	> 0.46	-0.73 – 0.65
Pengurusan	0.87 – 1.19	0.87 – 1.19	0.07	> 0.63	-0.59 – 0.42
Keusahawanan	0.81 – 1.27	0.81 – 1.27	0.07	> 0.71	-0.34 – 0.26
Tanggungjawab Sosial	0.92 – 1.14	0.93 – 1.14	0.06 – 0.07	> 0.57	-0.76 – 0.73

Menurut Linacre (2008) nilai peratus persetujuan antara penilai daripada pemerhatian yang dibuat adalah nilai peratus yang dipersetujui oleh seseorang penilai dengan penilai yang lain dalam menilai sesuatu item terhadap seseorang pelajar. Nilai jangkaan (expected)

melaporkan nilai persetujuan yang sepatutnya berdasarkan Model Pengukuran Rasch. Interpretasi bagi perbandingan nilai pemerhatian dan jangkaan ditunjukkan dalam Jadual 3.

JADUAL 3. Interpretasi Peratus Persetujuan antara Penilai Berdasarkan Nilai Peratus Pemerhatian (Observed) dan Jangkaan (Expected) Menurut Model PRPF

<i>Rasch Agreement</i>	Interpretasi
<i>Observed < Expected</i>	Menunjukkan percanggahan pendapat, biasa berlaku bagi penilai yang tidak dilatih.
<i>Observed ≈ Expected</i>	Penilai bertindak secara tersendiri. Perlu disahkan dengan nilai <i>fit</i> statistik
<i>Observed somewhat > expected</i>	Biasa bagi penilai yang terlatih. Latihan menitikberatkan persetujuan dengan yang lain tetapi <i>rating</i> memerlukan penilai bertindak secara bebas.
<i>Observed >> expected</i>	Penilai tidak bertindak secara bebas. Mungkin ada tekanan untuk bersetuju dengan yang lain.
<i>Observed > 90%</i>	Penilai adalah seperti “ <i>rating machines</i> ”. Perlu dikeluarkan sebagai faset dari model pengukuran

Sumber: Linacre (2008)

Hasil kajian adalah seperti di Jadual 4 menunjukkan bahawa nilai pemerhatian (*observed*) adalah melebihi dari nilai jangkaan (*expected*) dan menunjukkan bahawa penilai telah bertindak seperti penilai yang terlatih, namun begitu mereka tidak bergantung pada rakan penilai yang lain dalam membuat penilaian. Ini memenuhi kehendak penyelidik

dalam memastikan pensyarah penilai membuat penilaian yang adil.

Pengukuran yang melibatkan penilai memerlukan penilai bertindak secara konsisten. Namun begitu, kajian-kajian lepas menunjukkan sesuatu yang sukar untuk dilaksanakan (Lunz et al. 1989; Lumley & McNamara 1995;

JADUAL 4. Nilai Peratus Persetujuan di antara Penilai

Konstruk	Observed (%)	Expected (%)	Fit
Komunikasi	48.3	39.4	0.99
Kepemimpinan	48.1	40.9	0.99
Kerja Berpasukan	48.0	43.4	0.99
Pembelajaran Sepanjang Hayat & Pengurusan	49.4	44.2	1.01
Pemikiran Kritis & Penyelesaian Masalah	43.6	40.7	0.99
Etika & Moral Profesional	46.2	40.4	1.01
Pengurusan	45.6	39.9	0.97
Keusahawanan	43.4	37.6	0.96
Tanggungjawab Sosial	45.4	39.3	1.00

JADUAL 5. Nilai *Outfit Mean Square* Pensyarah Penilai bagi Sembilan Konstruk Kemahiran Generik

Konstruk	A	B	C	D	E	F	G	H	I
Penilai 1	1.13	1.08	1.12	1.18	1.02	1.02	1.12	0.95	0.71
Penilai 2	1.03	1.07	0.67	1.25	1.10	1.23	0.67	0.93	0.75
Penilai 3	0.84	0.99	1.47	0.98	0.96	0.87	1.47	0.94	0.81
Penilai 4	0.96	1.25	0.98	1.14	0.89	0.94	0.99	0.84	0.48
Penilai 5	1.34	1.21	1.05	0.53	1.37	0.29	1.05	1.38	1.19
Penilai 6	0.55	0.62	1.16	1.11	0.63	0.73	1.16	0.43	0.40
Penilai 7	0.95	1.05	1.36	1.32	1.12	1.17	1.36	0.99	0.97
Penilai 8	1.29	0.58	0.96	0.40	0.87	2.09	0.96	0.90	2.20
Penilai 9	1.28	1.24	1.02	1.08	1.22	1.07	1.02	1.24	1.19
Penilai 10	1.43	1.79	0.96	1.58	1.44	1.05	0.95	1.09	1.37
Penilai 11	0.76	0.91	0.96	0.83	0.91	0.78	0.95	0.91	0.74
Penilai 12	0.61	0.72	0.92	0.85	0.85	0.65	0.93	0.78	0.87
Penilai 13	0.94	0.86	1.11	0.68	0.81	0.93	1.11	0.82	0.98
Penilai 14	1.2	1.33	0.58	1.09	1.25	1.06	0.58	1.57	1.42
Penilai 15	1.16	1.08	0.58	0.99	1.13	1.01	0.58	1.23	0.91
Penilai 16	0.65	0.57	0.96	0.51	0.51	0.59	0.96	0.68	0.66
Penilai 17	0.63	0.57	0.98	0.52	0.53	0.65	0.98	0.60	0.56

Nota:

A : Komunikasi

B : Kepemimpinan

C : Kerja Berpasukan

D : Pembelajaran Sepanjang Hayat dan Pengurusan Maklumat

E : Pemikiran Kritis dan Penyelesaian Masalah

F : Etika dan Moral Profesional

G : Pengurusan

H : Keusahawanan

I : Tanggungjawab Sosial

Fitzpatrick et al. 1998). Konsistensi penilai boleh dilihat melalui nilai *infit* dan *outfit mean square* (Siti Rahayah et al. 2008c, 2008e; Linacre 2008). Jadual 5 menunjukkan *outfit mean square* bagi 17 pensyarah penilai bagi kesembilan konstruk kemahiran generik yang diukur. Nilai *outfit* yang diterima adalah di antara 0.5 dan 1.5 (Siti Rahayah et al. 2008a, 2008b; Linacre 2008). Hasil kajian menunjukkan terdapat beberapa penilai yang tidak begitu konsisten dalam menilai kemahiran generik pelajar. Penilai tersebut ialah penilai enam bagi keusahawanan dan tanggungjawab sosial, penilai lapan bagi pembelajaran sepanjang hayat dan pengurusan maklumat, etika dan moral profesional dan tanggungjawab sosial, penilai 10 bagi kepemimpinan dan pembelajaran sepanjang hayat dan pengurusan maklumat dan penilai 14 bagi pengurusan.

Menurut Matsuno (2006), penilai yang dikategorikan sebagai tidak konsisten juga adalah disebabkan terdapat percanggahan pendapat terhadap kualiti prestasi pelajar. Menurut Constable dan Andrich (1984), hubungan yang tinggi antara penilai mungkin dapat meningkatkan kebolehppercayaan tetapi akan mengurangkan kesahan. Setiap penilai mempunyai kaedah dan piawai tersendiri dalam membuat penilaian dan sekiranya latihan yang diberikan dapat menghapuskan perbezaan antara penilai, maka hasil pengukuran yang dibuat tidak akan dapat memberikan gambaran sebenar akan pandangan penilai terhadap pelajar yang dinilai. Sekiranya semua penilai memberikan penilaian yang sama, maka tidak ada gunanya menggunakan lebih daripada seorang penilai. Kaedah kedua yang boleh digunakan adalah dengan menyingkirkan penilai

yang tidak konsisten daripada analisis yang dilakukan. Program *Facets* yang digunakan dapat mengenal pasti penilai yang tidak konsisten dan penilaian yang dibuat oleh penilai tersebut tidak perlu diambil kira dalam mengukur kemahiran pelajar. Namun begitu, kaedah ini memerlukan lebih daripada seorang penilai menilai setiap pelajar dan bukanlah satu kaedah yang terbaik. Bagi situasi ujian yang berbeza, kedua-dua kaedah ini boleh diambil kira oleh pengguna-pengguna instrumen pengukuran yang melibatkan penilai dalam membuat pengukuran.

KESIMPULAN DAN IMPLIKASI

Kajian menunjukkan bahawa instrumen kemahiran generik berasaskan perspektif pensyarah ini mempunyai kebolehpercayaan dan kesahan yang sesuai dan memenuhi Model Pengukuran Rasch. Analisis penggunaan skala juga telah menunjukkan bilangan kategori yang digunakan adalah sesuai. Peratus persetujuan antara pensyarah juga memenuhi Model Pengukuran Rasch dan menunjukkan bahawa penilai bertindak secara bebas dan latihan yang diberikan membantu pensyarah dalam membuat penilaian.

Kajian ini juga menunjukkan bahawa penggunaan Model Pengukuran Rasch pelbagai faset dapat meningkatkan pengukuran apabila semua faset diteliti dan dianalisis. Item-item dan penilai yang *misfitting* boleh dikenal pasti agar penambahbaikan dapat dilakukan. Ini akan dapat menghasilkan satu tahap pengukuran yang objektif bagi instrumen yang memerlukan penilai dan pengukuran menjadi lebih bermakna dalam mengukur kemahiran generik pelajar.

Hasil kajian ini mempunyai tiga implikasi. Dari sudut teoretikal, kajian ini menyumbang secara langsung kepada perkembangan literatur pengukuran kemahiran generik di Malaysia, khususnya pengukuran kemahiran generik pelajar prauniversiti program matrikulasi menggunakan Model PRPF. Dari sudut metodologikal, pihak-pihak yang terbabit dengan pengukuran dan penilaian di institusi pengajian perlu peka bahawa kaedah pengukuran kemahiran generik pelajar secara sistematik pada zaman globalisasi ini wajar diberi perhatian serius kerana pengukuran dan penilaian kemahiran generik menggunakan model ini ternyata sekali adil dan objektif. Dari sudut pedagogikal, guru dan pensyarah perlu merancang dengan teliti segala program dan aktiviti agar pelajar boleh mempamerkan kemahiran generik mereka dengan jelas. Pengajaran dan pembelajaran hendaklah mengutamakan kaedah yang berpusatkan pelajar yang memberi peluang semaksimum mungkin untuk mereka belajar secara sendiri dan secara inkuiri berbantuan teknologi maklumat dan komunikasi. Dengan memastikan pelajar memahami kemahiran generik yang mereka telah miliki dan aspek yang perlu diperbaiki, baharulah pelajar dapat menjadi warga ekonomi global yang lebih produktif dan inovatif.

RUJUKAN

- Banno, E. 2008. Investigating an oral placement test for learners of Japanese as a second language. Tesis Ph.D. Temple University.
- Bond, T.G. & Fox, C.M. 2001. *Applying the Rasch Model: Fundamental Measurement in The Human Sciences*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Carnavale, A.P. Gainer, L.J. Meltzer, A.S. 1990. *Workplace Basic*. San Francisco, CA: Jossey-Bass.
- Cohen, R.J. & Swedlik, M.E. 2005. *Psychology testing and assessment: An introduction to test and measurement*. Ed. Ke-6. New York: McGraw-Hill.
- Constable, E. & Andrich, A. 1984. Inter-judge reliability: Is complete agreement among judges the ideal? Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans, LA. (Eric Document Reproduction Service No. ED243962).
- Downing, S.M. 2006. Twelve steps for effective test development. Dlm. *Handbook of test development*. Mahwah, New Jersey: Lawrence Erlbaum 3-26.
- Engelhard, G. 1994. Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of educational measurement* 31(2): 93-112.
- Engelhard, G. 1996. Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement* 33: 56-70.
- Engelhard, G. 1997. Constructing rater and task banks for performance assessments. *Journal of outcome measurement*. 1(1): 19-33.
- Engelhard, G. & Myford, C. 2003. *Monitoring faculty consultant performance in the advanced placement English literature and composition program with a many faceted Rasch model*. New York: College Entrance Examination Board.
- Ertmer, P. & Newby, T. 1993. Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly* 6(4): 50-72.
- Fitzpatrick, A. R., Ercikan, K., Yen, W. M. & Ferrara, S. 1998. The consistency between raters scoring in different test years. *Applied Measurement in Education* 11: 195-208.
- Gregory, R.J. 2007. *Psychological testing: History, principles and applications*. Ed. Ke-5. Boston: Pearson International.
- Grummon, P. 1993. Assessing educational outcomes: Trends and opportunities. In T. Bynum, J..
- Iramaneerat, C. 2007. A validity study of a communication skills assessment of medical residents. Tesis Ph.D. University of Illinois.
- Kearns, P. 2001. *Review of research: Generic skills for the new economy*. Adelaide. NCVER.
- Kondo-Brown, K. 2002. A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing* 19(1): 3-31.
- Linacre, J. M. 1994. Many-facet Rasch measurement. Chicago, IL: MESA Press.
- Linacre J. M. 1995. Categorical misfit statistics. *Rasch Measurement Transactions* 9(3): 450-451.
- Linacre J.M. 1997. Guideline for rating scale. Midwest Objective Measurement Seminar, Chicago, June 1997.

- Linacre, J.M. 2002. *What do Infit and Outfit, Mean-square and Standardized mean? Rasch Measurement Transactions* 16: 2.878.
- Linacre, J.M. 2005. Standard Errors: Means, Measures, Origins and Anchor Values. *Rasch Measurement Transactions* 19: 3 . 1030
- Linacre, J.M. 2006. Item Discrimination and Rasch-Andrich thresholds. *Rasch Measurement Transactions* 20: 1. 1054
- Linacre, J. M. 2008. *Facets Rasch Software Users Guide*. Chicago, IL: MESA Press.
- Ling, S.E., Siti Rahayah Ariffin, Saemah Rahman & Lai, K.L. 2010. Diversity in education using blended learning in Sarawak. *US-China Education Review*, USA. Feb 2010, Volume 7, No. 2 (Serial No. 63).
- Lumley, T. & McNamara, T. F. 1995. Rater characteristics and rater bias: Implications for training. *Language Testing* 12: 54-71.
- Lunz, M.E., Stahl, J.A., Wright, B.D. & Linacre, J.M. 1989. Variation among axaminers and protocols on oral examinations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. Eric Document TM 012988.
- Lunz, M.E., Wright, B.D. & Linacre, J.M. 1990. Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*. 3: 331-345.
- MacMillan, P. 2000. Classical, generalizability and multifaceted Rasch detection of interrater variability in large, sparse data sets. *Journal of Experimental Education* 68(2): 167-191.
- Matsuno, T. 2007. Self, peer and teacher assessment in Japanese university EFL writing classrooms. Thesis Ph.D. Temple University.
- Mayer Committee. 1992. *Putting General Education to Work: The Key Competencies Report*. AEC/MOVEET: Melbourne.
- McCurry, D. & Bryce, J. 1997. *The school based key competencies levels assessment project*. Final report. Department of Employment, Education, Training and Youth Affairs. Canberra.
- McCurry, D. & Bryce, J. 2000. *Victorian Board of Studies: Key competencies levels assessment trial*. Working paper 2. Victorian Curriculum and assessment authority, Melbourne.
- Myford, C. & Wolfe, E. 2002. When raters disagree, then what: examining a third-rating discrepancy solution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement* 3(3): 300-324.
- Popham, W.J. 1995. *Classroom assessment: What teachers need to know*. Boston, MA: Allyn and Bacon.
- Purya Baghaei. 2008. The Rasch Model as a Construct Validation Tool *Rasch Measurement Transactions* 22: 1 Summer 2008.
- Rodiah Idris, Siti Rahayah Ariffin & Noriah Mohd Ishak. 2008. Pemeriksaan ciri-ciri psikometrik instrumen kemahiran generik menggunakan pendekatan model Rasch. Prosiding Seminar Kebangsaan Jawatankuasa Penyelarasan Pendidikan Guru (JPPG 2008). Kementerian Pelajaran Malaysia.
- Rodiah Idris, Siti Rahayah Ariffin & Noriah Mohd Ishak. 2009. Application of Rasch Model in validating the construct of measurement for generic skills instrument for higher education (GESIHE). Conference of the Pacific Rim Objective Measurement Symposium (PROMS). 2009. 28 – 30 Julai, Hong Kong.
- Secretary's Commission on Archieving Necessary Skills (SCANS). 1991. *What work requires of schools*. Washington DC: United States Department of Labor.
- Sharifah Hapsah Syed Hasan Shahabudin. 2005. Kerangka Kelayakan Malaysia. Kuala Lumpur: Lembaga Akreditasi Negara.
- Shavelson, R. & Baxter, G. 1992. What we've learned about assessing hands-on science. *Educational Leadership* 49(8): 20-25.
- Siti Rahayah Ariffin. 2003. *Teori, Konsep dan Amalan Pengukuran dan Penilaian Pendidikan*. Bangi: Pusat Pembangunan Akademik. Universiti Kebangsaan Malaysia.
- Siti Rahayah Ariffin, Noriah Mohd Ishak, Abdul Ghafur Ahmad, Rosadah Majid, Norshidah Salleh & Rahaty Mohd Majzub. 2005. Manual penggunaan Malaysia Multiple Intelligences Checklist for Adult (MyMICA). Laporan Penyelidikan Pusat Penyelidikan dan Inovasi (PPPI). Universiti Kebangsaan Malaysia. 35-55.
- Siti Rahayah Ariffin. 2008. *Inovasi dalam Pengukuran dan Penilaian Pendidikan*. Bangi: Fakulti Pendidikan. Universiti Kebangsaan Malaysia.
- Siti Rahayah Ariffin, Rosadah Abdul Majid, Zolkefli Haron, Rodiah Idris, Shahrir Samsuri, Basri Hassan & Nor Azaheen Abdul Hamid. 2008a. Kemahiran Pemikiran Kritis dan Penyelesaian Masalah Pelajar-pelajar Sains di Malaysia. *Proceeding of Science and Mathematics Education Regional Conference 2008*. (SMEReC 2008). 15-20.
- Siti Rahayah Ariffin, Noriah Mohd Ishak, Rosadah Abdul Majid, Zolkefli Haron, Rodiah Idris, Nur 'Ashiqn Najmuddin, Nor Azaheen Abdul Hamid. Shahrir Samsuri & Basri Hassan. 2008b. Profil Kemahiran Pemikiran Kritis Remaja Malaysia. *Proceeding of Science and Mathematics Education Regional Conference 2008*. (SMEReC 2008). 20-29.
- Siti Rahayah Ariffin, Zolkefli Haron, Rodiah Idris, Nur 'Ashiqn Najmuddin, Shahrir Samsuri, Basri Hassan & Nor Azaheen. 2008c. Tahap penguasaan kemahiran pemikiran kritis dan penyelesaian masalah pelajar sains dan hubungannya dengan pencapaian. *Prosiding Seminar Pendidikan Kejuruteraan Dan Alam Bina*. (PEKA 2008). Fakulti Kejuruteraan dan Alambina. Universiti Kebangsaan Malaysia. 15-23.
- Siti Rahayah Ariffin, Noriah Mohd Ishak, Riza Atiq OK Rahmad, Abdul Ghafur Ahmad, Rodiah Idris, Nur 'Ashiqn Najmuddin, Benedict Clement D' Silva & Wan Ameran Wan Mat. 2008d. Assessing generic skills using Rasch model approach: A method for construct validity and reliability. *Proceeding for International Conference on Education Innovation (ICEI2008)*. 30-42.
- Siti Rahayah Ariffin, Ling, S.E., Saemah Rahman & Lai, K.L. 2008e. Perspectives on blended learning in Malaysia higher learning institutions: Need, readiness and future challenges. *Proceeding for ASAIHL international conference 2008*. Nonthaburi, Thailand. 5-12.
- Siti Rahayah Ariffin, Rosseni Ariffin dan Hafsa Mohamed Makki 2008h. Faktor Kontribusi Kecerdasan Pelbagai dalam Kalangan Pelajar Remaja (Factors Contributing To Multiple Intelligence among Teenagers). *Jurnal Pendidikan*. Vol. 35-46.
- Siti Rahayah Ariffin & Nor Azaheen Abdul Hamid. Profil Kemahiran Pemikiran Kritis Antara Pelajar Aliran Sains dan Bukan Sains. 2009. *Malaysian Education Deans' Council Journal*. Vol. 121-135.
- Weigle, S. C. 1998. Using FACETS to model rater training effects. *Language Testing* 15(2): 263-287.
- Wright B.D. 1995. Which standard error? *Rasch Measurement Transactions* 9(2): 436-7.

- Wright, B.D. & Masters, G.N. 1982. *Rating Scale Analysis*.
Chicago: MESA Press.
- Wright, B.D. & Stone, M.H. 1979. *Best Test Design*. Chicago, IL:
Mesa Press.

Untuk maklumat lanjut sila hubungi
Siti Rahayah Ariffin
Jabatan Asas Pendidikan
Fakulti Pendidikan
Universiti Kebangsaan Malaysia
43600 Bangi
Selangor D.E.