



# Analisis Kefungsian Skala Bagi Pembinaan Rubrik Pentaksiran Kompetensi Amali: Analisis Berdasarkan Model Pengukuran Rasch Pelbagai Facet

Azmanirah Ab Rahman<sup>1\*</sup>, Nurfirdawati Muhamad Hanafi<sup>2</sup>,  
Anizam Mohamed Yusof<sup>3</sup>, Halizah Awang<sup>4</sup>

<sup>1,2,3,4</sup>Fakulti Pendidikan Teknikal dan Vokasional,  
Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, MALAYSIA

\*Corresponding Author

DOI: <https://doi.org/10.30880/ojtp.2019.04.02.015>

Received 25<sup>th</sup> April 2019; Accepted 25<sup>th</sup> September 2019; Available online 30<sup>th</sup> September 2019

**Abstrak:** Pentaksiran prestasi seperti amali dalam pendidikan teknikal dan vokasional (PTV) sukar diukur kerana subjektif dan boleh menyebabkan bias. Penggunaan skala pemarkahan berbentuk rubrik dikatakan sesuai untuk mengukur kompetensi amali pelajar kerana telah menterjemahkan kriteria kualitatif kepada bentuk kuantitatif secara skala penarafan. Namun begitu pembinaan rubrik bukanlah perkara yang mudah terutamanya dalam menentukan skala penarafan. Oleh itu satu kajian telah dijalankan untuk menentukan kesesuaian skala penarafan yang digunakan dalam pembinaan rubrik kompetensi amali berdasarkan analisis kefungsian skala bersandarkan model Pengukuran Rasch Pelbagai Facet (MFRM). Kajian ini telah dijalankan di tiga buah Kolej Vokasional. Seramai sembilan pemeriksa telah memeriksa kompetensi amali pelajar elektronik dengan menggunakan rubrik empat mata yang telah dihasilkan ke atas 68 orang pelajar. Analisis kefungsian skala menunjukkan taburan frekuensi bagi setiap kategori adalah tidak teratur. Walau bagaimanapun, nilai fit pengukuran menghampiri 1. Selain dari itu juga nilai ambang dan nilai fit masih lagi berada di dalam nilai yang diterima. Secara keseluruhannya skala penarafan yang digunakan di dalam rubrik sesuai dengan sedikit penambahbaikan pada deskriptor yang paling ekstrem.

**Kata Kunci:** Kefungsian skala, model pengukuran rasch pelbagai facet

**Abstract:** Performance assessments such as laboratory practice in technical and vocational education (TVE) are difficult to measure because they are subjective in their characteristics, which can cause bias. The use of a rubric scoring scale is suitable for measuring students' practical competency as it could translate the qualitative criteria into quantitative forms on a grading scale. However the development of rubric is not easy especially in determine the rating scale. Therefore the study was conducted to determine the appropriateness of rating scale used in the construction of rubric based on the analysis of scale functionality of Many Facet Rasch Measurement model. Nine rater's have examined the competency performance of 68 students using developed four-point rating scale rubric consisting of 16 items. This study was conducted at three vocational colleges. Scale functionality analysis shows that the frequency distribution for each category is irregular. However, the fit value of the measurement approaches 1. In addition, the threshold value and the fit value are still within the accepted value. Overall the rating scale used in rubric suitable corresponds to a slight improvement on the most extreme descriptors.

**Keywords:** Scale functionality, many facet rasch measurement model



## 1. Pengenalan

Pentaksiran kompetensi bersifat terbuka dan subjektif. Tetapi bagaimana kita boleh membuat penilaian subjektif menjadi seobjektif yang mungkin serta penilaian yang konsisten antara satu pemeriksa dengan yang lain. Justeru satu standard pengukuran yang mengandungi kriteria dan prosedur yang jelas perlu dilaksanakan. Ini adalah kerana tugas prestasi tidak dapat ditaksir sekiranya tidak dipecahkan kepada tingkah laku spesifik dan ciri-ciri yang lebih kecil yang boleh dihakimi dengan mudah, serta boleh diperhatikan. Tanpa penggunaan kriteria yang jelas, penilaian sukar untuk dihakimi (Burke, 2001). Oleh itu penggunaan rubrik merupakan satu cara untuk menangani isu tersebut (Reddy, & Andrade 2010).

Rubrik merupakan alat pemarkahan yang sangat sesuai untuk mengukur tugas berbentuk autentik yang menilai proses kerja dan hasil kerja pelajar (Pickford & Brown 2007). Ciri-ciri rubrik yang direka bentuk sedemikian rupa menyediakan satu standard pemarkahan yang eksplisit kepada pengguna (Mabry 2004). Menurut Abdullah dan Shukur (2012) rubrik merupakan mekanisme yang komprehensif kepada pelajar untuk memenuhi standard yang telah ditentukan. Selain itu juga ia mampu memberi maklum balas kepada pelajar berkaitan pencapaian daripada pemeriksa (Allen 2011). Rubrik mempunyai potensi untuk membuat sumbangan besar kepada pengajaran berkualiti tetapi kita perlu membetulkan kelemahan yang membuat banyak rubrik hampir tidak bernilai.

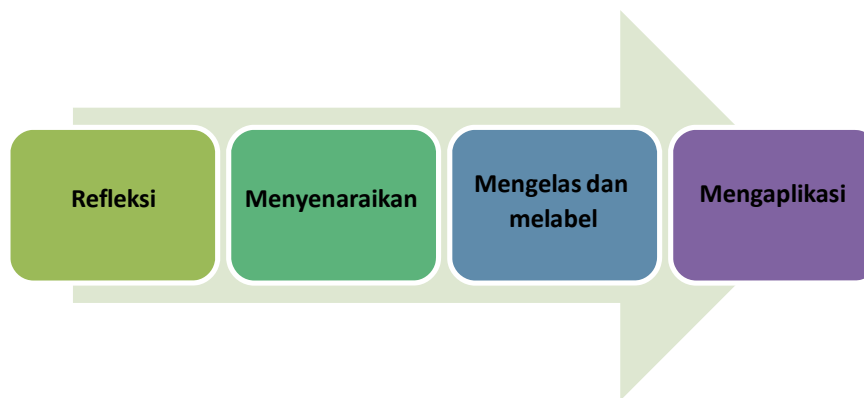
Bagaimana untuk menghasilkan rubrik yang berkualiti? Ramai yang anggap mudah untuk menghasilkan rubrik, namun menghasilkan rubrik yang berkualiti bukanlah perkara yang boleh dipandang remeh. Terdapat banyak proses yang perlu dilalui supaya rubrik yang dihasilkan bermakna dan memenuhi kehendak serta mempunyai definisi kualiti yang baik. Justeru tujuan artikel ini adalah untuk membincangkan tentang penentuan skala penarafan dalam penghasilan rubrik kompetensi amali untuk mengukur kompetensi penghasilan pelajar menghasilkan litar elektronik.

### 1.1 Rubrik Sebagai Alat Ukur

Rubrik merupakan panduan pemarkahan yang mengandungi kriteria dan digunakan oleh guru dan pelajar bagi menentukan kualiti tugas mencapai piawai yang ditetapkan (Mertler 2001; Goodrich 1996; Arter dan Tighe 2001). Ia merupakan mekanisme untuk menilai kualiti hasil kerja pelajar. Format rubrik mungkin berbeza tetapi elemen yang terdapat pada rubrik adalah konsisten. Menurut Popham (1997) rubrik terdiri daripada tiga ciri penting iaitu kriteria penilaian, definisi kualiti, dan strategi pemarkahan. Kriteria penilaian digunakan untuk membezakan tindak balas yang diterima daripada tindak balas yang tidak diterima. Kriteria adalah berbeza-beza bergantung kepada kemahiran yang hendak dinilai. Definisi kualiti menerangkan perbezaan kualitatif dalam jawapan pelajar. Rubrik perlu memberikan penerangan yang berasingan bagi setiap peringkat kualitatif. Jika terdapat empat aras yang berbeza, rubrik haruslah mempunyai definisi yang berbeza bagi setiap peringkat. Manakala strategi pemarkahan terdiri daripada holistik atau analitik.

Proses mereka bentuk rubrik bukanlah perkara yang mudah (Goodrich, 1996). Beberapa perkara perlu diberi perhatian. Terdapat banyak kaedah yang boleh dilaksanakan oleh pembina rubrik seperti yang dicadangkan oleh Arter dan McTighe (2001), Mertler (2001), Marie (2004), Goodrich (1996), Reddy (2010), Burke (2011) dan Steven & Levi (2013). Walaubagaimanapun dalam kajian ini pengkaji memilih menggunakan model pembinaan rubrik oleh Stevens dan Levi (2013) seperti yang digambarkan pada Rajah 1. Model ini sangat sesuai kerana mempunyai empat ciri penting pembinaan rubrik iaitu refleksi, menyenaraikan, mengelaskan dan melabel, dan mengaplikasi.

Proses refleksi adalah untuk menentukan objektif pengajaran selari dengan pentaksiran dan apakah yang ingin dicapai, mengapa tugas tertentu diwujudkan dan apa yang berlaku kali terakhir tugas itu diberikan dan jangkaan untuk pelajar tersebut (Mertler 2001; Marie 2004; Reddy 2010). Proses menyenaraikan pula merupakan proses menentukan kriteria penilaian adalah melalui kaedah sumbang saran (Goodrich 1996; Mertler 2001), pengalaman guru dan sorotan kajian serta adaptasi dari rubrik sedia ada (Arter & Mc Tighe 2001), model tugas yang lepas dengan melibatkan pelajar secara aktif untuk menentukan kriteria (Mertler 2001) serta menentukan idea utama (Burke 2011). Proses ketiga iaitu mengelaskan dan melabel adalah proses untuk melabelkan jangkaan prestasi yang serupa. Jangkaan prestasi yang serupa adalah kumpulan di bawah tajuk organisasi. Label ini menjadi kriteria bagi rubrik. Kriteria yang disenaraikan dikumpulkan berdasarkan kumpulan sama ada baik atau lemah serta ditentukan juga gred pertengahan (Goodrich 1996; Mertler 2001). Akhir sekali ialah proses mengaplikasi merupakan proses pengumpulan maklumat yang diorganisasikan dari senarai dan kumpulan dimasukkan ke dalam bentuk grid. Pada ketika ini bentuk rubrik boleh menjadi panduan untuk pemarkahan. Pelajar boleh menggunakan rubrik sebagai penilaian sendiri dan juga rakan (Goodrich 1996; Marie 2004).



Rajah 1 - Proses pembinaan rubrik (Stevens dan Levi, 2013)

## 1.2 Penentuan Skala Penarafan Dalam Penghasilan Rubrik

Selain dari model pembinaan rubrik, skala penarafan memainkan peranan penting dalam pembinaan rubrik. Skala penarafan akan menyalurkan cara bagaimana pemeriksa menilai dengan konkrit prestasi sampel dan akhirnya memberikan markah kepada calon (Eckes 2008). Namun kadangkala responden gagal untuk bertindak balas seperti yang dikehendaki oleh pembina konstruk. Oleh itu skala pemeringkatan perlulah didefinisikan dengan baik, *mutually exclusive*, *univocal* dan menyeluruh semasa membuat kategori skala (Linacre 1999).

Tidak ada jumlah yang khusus untuk bilangan skala penarafan bagi rubrik. Walau bagaimanapun panjang skala turut memberi kesan kepada pemeriksa. Semakin banyak skala penarafan maka semakin sukar untuk membezakan di antara satu skala dengan skala lain. Skala penarafan yang terlalu banyak menyukarkan pemeriksa untuk membezakan pernyataan bagi setiap skala malah memori manusia agak sukar untuk mengingat setiap pernyataan pada setiap skala (Penny et al. 2000). Di samping itu juga akan menjadi satu beban kepada pemeriksa. Skala yang terlalu sedikit pula akan menyebabkan kurangnya variasi di antara satu elemen dengan elemen yang lain (Schaefer 2008). Seperti Thompson et.al (2013) mencadangkan penggunaan skala yang sedikit iaitu 3 atau 4. Tetapi bagi Coffman (1971) pula menyatakan skala yang sedikit pula tidak dapat membezakan kebolehan pelajar dengan jelas dan akan menghasilkan kebolehpercayaan yang rendah. Banyak hasil kajian membuat kesimpulan yang berbeza berkaitan skala pemeringkatan yang optimal dalam kes kebolehpercayaan seperti (Coffman 1971; Longford 1994; Myford 2002; Thompson 2013).

Terdapat lapan panduan yang digariskan oleh Linacre (1999) untuk memastikan skala pemeringkatan yang digunakan berfungsi dengan baik berdasarkan model Many Facet Rasch Measurement (MFRM) iaitu (i) Sekurang kurangnya 10 pemerhatian bagi setiap kategori (ii) pemerhatian taburan berkala (iii) pengukuran purata meningkat secara senada dengan skala kategori (iv) Outfit MNSQ kurang dari 2.0. Untuk model Rasch statistik MNSQ fit telah ditentukan pada nilai 1.0. Nilai yang lebih besar dari 1.5 menunjukkan rawak yang tidak dapat dijelaskan. Nilai yang lebih besar dari 2.0 pula menunjukkan lebih banyak hingar yang tidak dapat dijelaskan. MNSQ yang besar menunjukkan segmen pada data tidak menyokong pengukuran yang bermakna. Kes ini berlaku mungkin disebabkan oleh kurangnya penggunaan skala pada kategori skala tersebut (v) *step calibration advance* (vi) Kedudukan menandakan pengukuran, dan pengukuran menandakan kedudukan (vii) *step difficulties advance by at least 1.4 logit* (viii) *step difficulties advance by less than 5.0 logit*.

Tujuan garis panduan yang disediakan bertujuan membantu penyelidik menambah baik kefungsi skala. Bukan semua panduan yang disediakan adalah sesuai kepada mana-mana analisis data tertentu. Walau bagaimanapun empat cara yang telah digunakan untuk menentukan kefungsi skala dalam kajian ini iaitu : (i) Frekuensi kategori (ii) Pengukuran purata (iii) Nilai ambang dan (iv) Nilai fit kategori (Bond, & Fox 2007). Kesemua analisis ini boleh diperolehi berdasarkan statistik kategori yang dijana oleh program FACET.

## 2. Metodologi

Kajian pembinaan rubrik bermula dengan pendekatan kualitatif dan diikuti pendekatan kuantitatif. Pendekatan kualitatif adalah secara temu bual dan juga analisis dokumen untuk mendapatkan kriteria dan definisi kualiti bagi rubrik kompetensi amali. Seramai tiga orang guru pakar telah ditemu bual secara mendalam untuk tujuan tersebut. Selain itu dokumen sedia ada dari sembilan KV juga dikumpul untuk mengenal pasti kriteria utama dalam rubrik yang dibangunkan. Kriteria yang diperolehi telah dijadikan dalam format rubrik. Rubrik yang dibangunkan merupakan rubrik analitik dengan empat skala penarafan. Skala yang dipilih tidak terlalu panjang dan tidak terlalu pendek supaya tidak wujud unsur bias semasa pemeriksaan. Rubrik yang telah siap disemak telah melalui kesahan kandungan dan juga kesahan konstruk. Kesahan telah dibuat oleh sebelas orang pakar yang berpengalaman dalam bidang pengukuran, guru yang mengajar bidang serta panel industri.

Pendekatan kuantitatif pula dilaksanakan untuk mengenal pasti adakah kategori skala bagi rubrik kompetensi amali yang telah dihasilkan berfungsi dengan efektif. Analisis data telah dilakukan dengan menggunakan program FACET bersandarkan model Many Facet Rasch Measurement (MFRM). Seramai sembilan orang pemeriksa telah dipilih secara

bertujuan untuk menggunakan rubrik yang telah dibangunkan dengan memeriksa kompetensi terhadap 68 pelajar dalam amali mereka bentuk litar elektronik. Pemeriksa telah diberi penerangan dan latihan menggunakan rubrik. Tujuan latihan penggunaan rubrik dijalankan adalah untuk memberi kefahaman yang jelas kepada pemeriksa mengenai definisi kualiti bagi setiap skala pemeringkatan supaya mereka mempunyai kefahaman yang sama.

### 3. Dapatan Kajian

Analisis bagi skala penarafan bertujuan untuk menentukan kesesuaian penggunaan skala di dalam instrumen pentaksiran kompetensi amali. Terdapat empat perkara yang dianalisis iaitu (i) Frekuensi kategori (ii) Purata pengukuran (iii) Nilai fit kategori dan (iv) Nilai ambang dengan berpandukan keluk kebarangkalian (probability curve).

Analisis frekuensi kategori ditentukan dengan memeriksa setiap kategori secara deskriptif untuk setiap maklum balas yang diberikan oleh pemeriksa. Merujuk Jadual 1, lajur pertama adalah skala kategori, lajur kedua menunjukkan kekerapan pemeriksa menanda setiap skala manakala lajur ketiga menunjukkan peratusannya. Dari jadual tersebut menunjukkan hanya 3% menggunakan skala 1, 21% menggunakan skala 2, 47% menggunakan skala 3 dan 28% menggunakan skala 4. Peratusan penggunaan skala 3 adalah tertinggi berbanding skala lain. Nilai peratusan menunjukkan bahawa taburan bagi setiap skala kategori adalah tidak teratur.

Jadual 1- Struktur skala penarafan

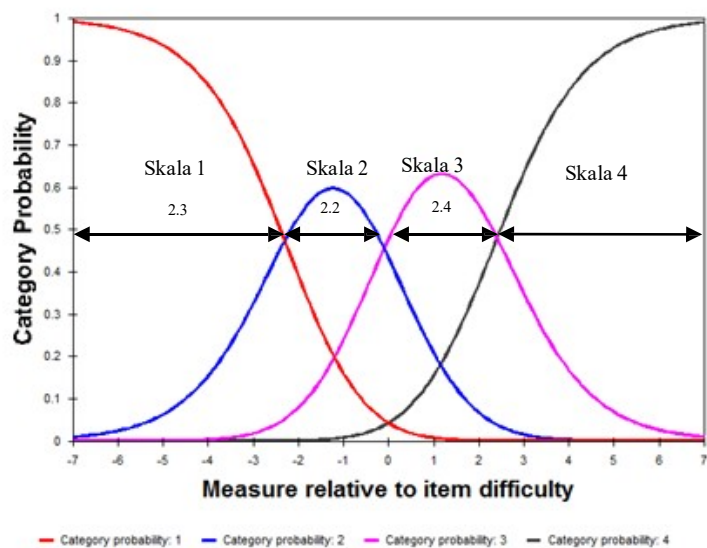
Skala Kategori	Bilangan frekuensi	Peratus pengguna	Purata Ukuran	Outfit MNSQ	Nilai ambang Rasch
1	9	3	-0.99	0	Tia
2	7	2	0.26	1	-
3	1549	4	1.23	1	-
4	93	28	2.58	1.0	2.4

Analisis kedua ialah menentukan purata ukuran. Purata pengukuran perlu meningkat secara senada (monotonically) dengan kategori skala. Ini bermakna semakin meningkat kategori skala, maka purata ukuran juga akan meningkat (Linacre 1999; Bond, & Fox 2007). Berdasarkan dapatan yang diperoleh pada skala 1 nilai bagi ukuran purata ialah -0.99. Pada skala 2 pula ukuran purata ialah 0.26. Pada skala 3, ukuran purata ialah 1.23 manakala pada skala 4 pula ukuran purata ialah 2.58. Berdasarkan dapatan yang diperoleh, nilai bagi purata pengukuran meningkat secara senada dengan kategori iaitu dari -0.99 hingga 2.58.

Indikator lain bagi menentukan kefungsiian bagi skala pemeringkatan ialah dengan melihat statistik *outfit* MNSQ. Analisis fit merupakan perbezaan antara ukuran kebolehan pelajar secara purata berbanding ukuran jangkaan kebolehan pelajar secara. Semakin besar perbezaan, maka semakin besar nilai *outfit* MNSQ. Namun begitu nilai *outfit* MNSQ secara statistik tidak boleh melebihi 2.0 Nilai yang lebih besar dari 2 menunjukkan terdapatnya hingar di dalam kategori yang diukur semasa proses pengukuran (Linacre 1999). Berdasarkan dapatan yang diperoleh, nilai *outfit* MNSQ adalah antara 0.8 hingga 1.1. Nilai ini masih berada dalam julat yang dibenarkan iaitu tidak melebihi 2.

Analisis keempat bagi menentukan kefungsiian skala penarafan ialah dengan menentukan nilai ambang Rasch. Nilai ambang juga perlu meningkat secara berkadar terus dengan kategori. Namun begitu perlu berada di antara julat 1.4 hingga 5 (Linacre 1999). Sekiranya nilai ambang kurang dari 1.4 skala tersebut perlu digabungkan dan jika nilai ambang melebihi 5, skala tersebut perlulah dipisahkan (Bond, & Fox 2007). Rajah 1 menunjukkan lengkung kebarangkalian untuk empat skala pemeringkatan yang digunakan oleh pemeriksa semasa memberi markah kepada pelajar bagi rubrik reka bentuk. Paksi 'x' mewakili skala kebolehan pelajar (examinee proficiency) dan paksi 'y' mewakili kebarangkalian daripada yang dinilai dalam setiap kategori. Satu lengkung mewakili satu kategori. Titik persilangan menunjukkan ambang Rasch Andrich. Titik persilangan bagi skala 1 dan skala 2 ialah -2.3. Titik persilangan antara skala 2 dan skala 3 ialah -0.1. Titik persilangan antara skala 3 dan 4 pula ialah 2.4. Perbezaan antara titik persilangan tersebut menghasilkan nilai dalam julat yang dibenarkan seperti yang ditunjukkan pada Rajah 2.

Secara keseluruhannya, analisis skala penarafan yang diperoleh menunjukkan taburan frekuensi bagi setiap kategori adalah tidak teratur. Walau bagaimanapun, nilai fit pengukuran menghampiri 1. Selain dari itu juga nilai ambang dan nilai fit masih lagi berada di dalam nilai yang diterima. Oleh itu skala penarafan yang digunakan dalam rubrik kompetensi adalah sesuai.



Rajah 2- LengkuK Kebarangkalian

#### 4. Perbincangan

Hasil dapatan mendapati skala penarafan empat mata yang digunakan dalam rubrik adalah sesuai dari tiga aspek iaitu pengukuran purata yang meningkat senada dengan kategori, nilai ambang di antara 1.4 hingga 5 dan nilai fit kategori kurang dari 2.0 (Bond, & Fox 2007). Walau bagaimanapun frekuensi kategori menunjukkan peratusan penggunaan skala adalah tidak teratur bagi setiap kategori. Hal ini berlaku kerana pemeriksa sangat cenderung menggunakan skala pertengahan iaitu 3. Ini berlaku mungkin kerana pemeriksa baru pertama kali menggunakan rubrik maka sukar untuk membezakan skala terutamanya skala 3 dan skala 4 walaupun telah diberi latihan awal penggunaan rubrik. Oleh itu ramai pemeriksa yang cenderung memberikan skor yang tinggi kepada pelajar sehingga terdedah kepada kesan kecenderungan memusat. Kecenderungan memusat ialah kecenderungan pemeriksa menggunakan skala pertengahan (Eckes 2011; Myford, & Wolfe 2004).

Pengkaji juga telah membuat andaian bahawa skala 4 pada rubrik terlalu mudah untuk dicapai sehinggakan ramai pemeriksa yang memilih memberi skor pada skala tersebut. Dapatan ini selari dengan pendapat Myford, & Wolfe (2004) bahawa ralat bukan sahaja berpunca dari pemeriksa jika ramai pemeriksa yang cenderung menggunakan skala tertentu, mungkin juga masalah dengan skala penarafan yang digunakan di dalam instrumen. Oleh itu pengkaji perlu menambah baik skala 4 dengan menggunakan definisi kualiti yang benar-benar boleh dibezakan dari skala 3. Selain itu pengkaji juga memberi perhatian terhadap skala-skala lain. Jadual 2 menunjukkan contoh rubrik sebelum penambahbaikan dan Jadual 3 Berikut merupakan penambahbaikan terhadap deskriptor pada setiap skala penarafan bagi kriteria rekabentuk litar. Jika dilihat pada Jadual 2, skala 4 dan skala 3 tidak dapat dibezakan dengan jelas oleh pemeriksa kerana pernyataan pada descriptor adalah hampir sama. Begitu juga pernyataan pada skala 3 dan skala 2. Setelah analisis kefungsiian dibuat pengkaji telah membuat penambahbaikan terhadap skala 2, skala 3 dan skala 4 seperti pada Jadual 3.

Selain itu juga frekuensi kategori yang tidak seimbang adalah disebabkan oleh pengkaji yang ada bersama-sama dengan pemeriksa semasa pemeriksa melaksanakan pemeriksaan ke atas pelajar. Pemerhatian oleh pengkaji telah menjadikan pemeriksa cenderung untuk menggunakan skala yang tinggi semasa memberikan markah kepada pelajar. Hal ini adalah selari dengan kajian Knoch et al. (2007) di mana pemerhatian yang dijalankan kepada pemeriksa menjadikan pemeriksa memberikan markah jalan selamat (play it safe strategy). Pengkaji juga mengambil pendekatan Farrokhi et al. (2011) dan Myford, & Wolfe (2004) untuk memberi maklum balas secara individu kepada pemeriksa supaya penerangan tentang skala pemeringkatan dijelaskan dengan lebih teliti dan jelas supaya mereka boleh mencapai kesepakatan yang sama dan mempunyai konsistensi di dalam pemeriksaan.

Jadual 2-Contoh rubrik sebelum penambahbaikan

Kriteria	1	2	3	4
Proses Rekabentuk Litar	Tidak dapat merekabentuk litar dengan baik atau mengambil idea orang	Merekabentuk litar dengan bantuan guru dan rakan.	Merekabentuk litar mengikut kreativiti sendiri hasil dari perbincangan dengan	Merekabentuk litar mengikut kreativiti dan idea sendiri.

**Jadual 3-Contoh rubrik selepas penambahbaikan**

<b>Kriteria</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
Idea rekabentuk litar	(i) Tidak dapat merekabentuk litar dengan baik dan memerlukan bantuan guru sepenuhnya. (ii) Tidak mengambil kira aspek kedudukan komponen, ruang komponen, jarak antara komponen, serta kekutuban komponen.	(i) Kurang berkebolehan untuk merekabentuk litar menggunakan kreativiti sendiri namun selepas beberapa kali perbincangan dengan guru akhirnya pelajar berupaya menghasilkan rekabentuk litar. (ii) Kurang mengambil kira aspek kedudukan komponen, ruang	(i) Boleh merekabentuk litar dengan baik mengikut kreativiti sendiri selepas mendapat sedikit idea hasil dari perbincangan dengan guru. (ii) Menggunakan idea yang baik semasa rekabentuk litar dengan mengambil kira aspek kedudukan komponen, ruang komponen, jarak	(i) Boleh merekabentuk litar dengan sangat baik mengikut kreativiti sendiri tanpa bantuan guru dan rakan. (ii) Menggunakan idea yang sangat kreatif dan inovatif semasa rekabentuk litar dengan mengambil kira aspek-aspek yang penting seperti kedudukan komponen, ruang komponen, jarak antara komponen

## 5. Kesimpulan

Berdasarkan dapatan yang diperoleh, skala penarafan yang digunakan dalam rubrik kompetensi amali yang dihasilkan adalah sesuai berdasarkan analisis menggunakan FACET. Skala empat mata yang dibina adalah sesuai berdasarkan analisis kefungsi skala. Analisis frekuensi bagi setiap kategori melebihi 10 pemerhatian, purata pengukuran pula meningkat berkadar terus dengan peningkatan skala kategori, semua skala fit dengan model Rasch dan nilai jarak ambang berada di antara 1.4 hingga 5. Terdapat perkara penting yang perlu diambil perhatian semasa pembinaan rubrik ialah pembina rubrik perlu lebih sensitif terhadap deskriptor bagi setiap skala yang digunakan supaya betul-betul dapat membezakan di antara satu skala dengan skala lain. Selain itu latihan tentang penggunaan rubrik mengenai skala penarafan yang digunakan juga perlu dijelaskan dengan lebih teliti supaya pemeriksa jelas dengan setiap descriptor yang digunakan dan mereka boleh mencapai kesepakatan yang sama dan mempunyai konsistensi dalam membuat pemeriksaan.

## Penghargaan

Penulis ingin merakamkan ribuan terima kasih kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam usaha menjayakan kajian ini.

## References

- Abdullah, E. & Shukor, A. A. A. 2012. *Pentaksiran Prestasi dan Pentaksiran Rujukan Standard dalam Bilik Darjah*. Universiti Pendidikan Sultan Idris Tanjong Malim.
- Airasian, P. W. 1991. *Classroom assessment*. New York : McGraw-Hill.
- Arter, J. & Chappius, J. 2007. *Creating and Recognizing Quality Rubric*. Upper Saddle River, New Jersey : Pearson.
- Arter, J. & McTighe, J. 2001. *Scoring Rubric in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Corwin Press, INC.
- Bond, T. G. & Fox, C. M. 2007a. *Applying the Rasch Model : Fundamental Measurement in the Human Sciences* hlm.2nd Editio. Lawrence Erlbaum Associates Publishers.
- Bond, T. G. & Fox, C. M. 2007b. *Applying the Rasch Model : Fundamental Measurement in the Human Sciences* hlm.2nd Edisi . Lawrence Erlbaum Associates Publishers.
- Burke, K. 2001. *From Standards to Rubric in Six Steps*. A Sage Company.
- Coffman, W. E.,(1971). On the Reliability of Ratings of Essay Examinations in English Research in the Teaching of English. Vol. 5, No. 1 (Spring, 1971), pp. 24-36.

- Farrokhi, F., Esfandiari, R. & Dalili, M. V. 2011. Applying the Many-Facet Rasch Model to Detect Centrality in Self-Assessment, Peer-Assessment and Teacher assessment. *World Applied Sciences Journal*, 15, 70–77.
- Goodrich, H. 1996. Understanding rubrics. *Educational Leadership*, 54(4), 14–17.
- Knoch, U., Read, J. & von Randow, J. 2007. Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26–43. doi:10.1016/j.asw.2007.04.001
- Linacre, J. M. 1999. Investigating rating scale category utility. *Journal of outcome measurement*, 3(2), 103–122. Mertler, C. a. 2001. Designing Scoring Rubrics for Your Classroom. *Practical Assessment Research Evaluation*, 7(25), 1–10. <http://pareonline.net/getvn.asp?v=7&n=25> June 12, 2014]
- Myford, C. M. & Wolfe, E. W. 2003. Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I. *Journal of Applied Measurement*, 5(2), 189–227.
- Myford, C. M. & Wolfe, E. W. 2004. Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part II. *Journal of Applied Measurement*, 5(2), 189–227.
- Pickford, R. & Brown, S. 2007. *Assessing Skills and Practice*. London and New York: Routledge Taylor and Francis Group.
- Popham, W. J. 1997. What's Wrong and What's right with Rubrics. *Educational Leadership*, 55(2), 72–75.
- Reddy, Y. M. & Andrade, H. 2010. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. doi:10.1080/02602930902862859
- Schaefer, E. 2008. Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465–493. doi:10.1177/0265532208094273
- Steven, D. D. & Levi, A. J. 2013. *Introduction to Rubrics* hlm.2nd Editio. Stylus Publishing.
- Thompson, M. K., Clemmensen, L. K. H., & Ahn, B. (2013). The Effect of Rubric Rating Scale on the Evaluation of Engineering Design Projects. *International Journal of Engineering Education*, 29(6), 1490–1502.