

## Sistem Pemarkahan Berasaskan Pemberat dan Persentil Taburan (Assessment System Based on Weight and Percentile of Distribution)

HAMIZUN BIN ISMAIL\*, AZMIN SHAM RAMBELY & ROKIAH ROZITA AHMAD

### ABSTRACT

*Kaedah pemarkahan dan penggredan sering dibincangkan dalam aspek penilaian pelajar. Kertas kerja ini bertujuan untuk membincangkan kaedah pemarkahan dan penggredan menggunakan persentil taburan markah. Kaedah pemarkahan dengan menggunakan min dan songsangan varians markah bagi setiap soalan dalam suatu peperiksaan sebagai pemberat kepada soalan berkenaan dibincangkan. Dalam membincangkan penggredan menggunakan persentil taburan markah, persentil boleh ditentukan dengan menganggar taburan bagi markah peperiksaan. Penganggaran dan penentuan persentil bagi dua taburan iaitu taburan normal dan taburan gamma digunakan. Perbandingan dilakukan terhadap keputusan bagi beberapa senario yang diperoleh. Keputusan menunjukkan bahawa kaedah pemarkahan berpemberat yang menggunakan pemberat yang berasaskan kepada tahap kesukaran soalan boleh digunakan untuk mernghasilkan peringkat gred yang lebih jelas dalam kalangan pelajar.*

*Kata kunci: Penggredan; penilaian; tahap kesukaran*

### ABSTRAK

*Assessment and grading methods are two important aspects in student evaluation. This paper proposes a method of assessing and grading using percentile of the distribution of examination scores. Assessment methods using the mean and the inverse of variance from the distribution in an examination as weights were discussed. In grading using percentile of the distribution of scores, percentiles can be determined by estimating the distribution of exam scores. Estimation and determination of percentile of the distribution of two distributions, namely normal and gamma distributions were used. Comparisons of results obtained under different scenarios were presented. The result showed that the weighted scoring method that gives weights based on the level of difficulty of the questions can be used to formulate a comprehensible grade distinction among students.*

*Keywords: Assessment; grading; level of difficulty*

### PENGENALAN

Penggredan dan pemarkahan bukanlah satu masalah besar. Pensyarah umumnya mampu membuat penilaian yang bijak mengikut pertimbangan serta pengalamannya. Walau bagaimanapun, perubahan corak dalam taburan pencapaian pelajar kadang kala menimbulkan masalah. Misalnya, apabila terlalu ramai pelajar yang gagal sesuatu kursus berbagai masalah yang timbul. Beberapa cadangan telah diutarakan untuk menangani isu ini, antaranya adalah menggunakan kaedah penilaian Sismuka iaitu menilai pelajar berdasarkan pencapaian keseluruhan semua pelajar di dalam kumpulan tersebut menggunakan suatu lengkung, Indeks Kesukaran, Indeks Diskriminasi dan lain-lain. Walau bagaimanapun, sesetengah pihak tidak bersetuju menggunakan kaedah Sismuka ini kerana ia mungkin bersifat agak tidak adil dalam penilaian pelajar dalam semester atau sesi yang berbeza. Oleh itu, beberapa kajian telah dilakukan dengan mengambil kira tahap kesukaran soalan yang diajukan menggunakan analisis item yang merangkumi indeks diskriminasi dan indeks kesukaran (Caulkins et al. 1996; Jandaghi 2010; Jandaghi & Shaterian

2008; Johari et al. 2011; Sarina et al. 2006; Sim & Rasiah 2006; Zamri et al. 2008; Zurawski 1998).

Biasanya indeks kesukaran dan indeks diskriminasi ini dilakukan terhadap soalan berbentuk objektif (Sarina et al. 2006; Sim & Rasiah 2006) tetapi Sarina et al. (2006) menggunakan rangkaian neural rambatan balik dalam mengelaskan tahap kesukaran soalan. Beberapa kaedah digunakan untuk mengukur tahap kesukaran soalan berbentuk subjektif, antaranya menggunakan pekali kesukaran dan pekali diskriminasi (Jandaghi 2010; Jandaghi & Shaterian 2008), Model Rasch (Zamri et al. 2008) dan purata skor (Juridah et al. 2011).

Terdapat beberapa kajian yang menarik berkaitan isu penilaian dan pemarkahan ini. Penny dan Grover (1996) dalam kajian mereka merumuskan bahawa pemeriksa sering terdorong kepada unsur subjektif bila memeriksa. Sebagai jalan penyelesaian, mereka menekankan keperluan terhadap pemeriksa kedua atau pemeriksa luar. Kniveton (1996) dalam kajiannya tentang penilaian berterusan menyimpulkan bahawa walaupun penilaian yang telus boleh menimbulkan tekanan, ia memberikan pelajar

motivasi, berguna untuk menjamin kehadiran kuliah yang baik, serta sebagai cara yang berkesan untuk memastikan pelajar-pelajar sentiasa mengikuti perancangan sesuatu kursus. Birenbaum (1997) pula, dalam kajiannya tentang strategi pembelajaran pelajar kejuruteraan dan pelajar pendidikan, mengenal pasti tiga faktor utama dalam kaedah pembelajaran: motivasi (contohnya matlamat, jangkaan dan harapan), strategi (contohnya penekanan kepada latihan dan rujukan) dan pengurusan sumber (contohnya masa, perbincangan, pertemuan dengan tutor/pensyarah, persekitaran). Dalam kajian tersebut, Birenbaum mendapati perbezaan strategi dan penekanan yang ketara dalam kalangan pelajar, tetapi perbezaan antara bidang tidak begitu ketara. Beliau juga mendapati perbezaan tersebut memberi kesan kepada pencapaian. Dalam konteks yang lain, Young dan Warrington (1996) dan Duru-Bellat (1996) pula membincangkan tentang pengaruh faktor-faktor sosial terhadap pencapaian akademik, serta menggariskan beberapa cadangan untuk memperbaiki keadaan.

Oleh itu, kertas ini bertujuan untuk membincangkan dua perkara, iaitu pemarkahan dengan menggunakan min dan songsangan varians dan penggedan menggunakan persentil taburan markah. Beberapa alternatif berkaitan kaedah pemarkahan dan penggedan tersebut diutarakan, dan keputusan-keputusan yang terhasil daripada kaedah-kaedah tersebut dibandingkan.

#### PEMARKAHAN BERPEMBERAT

Lazimnya markah penuh yang diumpukkan kepada setiap soalan dalam suatu peperiksaan ditentukan sama ada dengan memberikan markah berpeMBERAT sama nilai bagi setiap soalan atau sebaliknya. Jika markah berpeMBERAT sama nilai digunakan, ini mengimplikasikan yang tahap kesukaran dan masa yang diperlukan untuk menyelesaikan soalan-soalan berkenaan tidak diambil kira dalam pengumpulan markah, dan/atau tahap kesukaran dan masa yang diperlukan untuk menyelesaikan soalan-soalan berkenaan diandaikan lebih kurang setara bagi setiap soalan. Mengumpukkan markah yang berbeza bagi setiap soalan merupakan satu cara untuk mengambil kira unsur-unsur kesukaran dan masa ke dalam pemarkahan.

Pengagihan markah dengan kaedah sedemikian berlandaskan tanggapan (*a priori*) semata-mata, iaitu tanggapan yang tidak tepat akan menghasilkan keputusan pemarkahan yang kurang wajar. Lebih wajar jika pengagihan markah dilakukan dengan memberi pemberat kepada setiap soalan dalam suatu peperiksaan itu berdasarkan tahap kesukaran serta masa yang diperlukan untuk menyelesaikan soalan-soalan berkenaan (selepas ini akan dirujuk sebagai tahap kesukaran) yang disukat dari keputusan awal peperiksaan tersebut. Secara kasarnya, sebagai langkah pertama, markah awal peperiksaan ditentukan berdasarkan pemarkahan berpeMBERAT *a priori*. Tahap kesukaran setiap soalan kemudiannya disukat dengan meneliti markah bagi setiap soalan. Sukatan

kesukaran bagi setiap soalan dijemakan sebagai pemberat bagi soalan tersebut. Akhirnya, markah muktamad dikira berdasarkan pemberat berkenaan.

#### MIN MARKAH AWAL SEBAGAI SUKATAN KESUKARAN

Katakan untuk  $(i,j) \in -\{(1,2,3, \dots, J), (1,2,3, \dots, J)\}$ ,  $m_{i,j}$  merupakan markah awal yang diperoleh pelajar  $i$  bagi soalan ke  $j$ , dan  $m_i$  merupakan markah penuh awal bagi pelajar  $i$ , dengan markah penuh yang diumpukkan untuk soalan  $j$ ,  $p_{j,p}$  bagi  $j = 1,2,3, \dots, J$  (lazimnya  $p_{j,J} = 100$ ). Min markah bagi soalan  $j$ :

$$\bar{m}_j = \frac{1}{I} \sum_i m_{i,j}, \quad (1)$$

boleh digunakan bagi menyukat kesukaran. Misalnya, anggapan soalan ke  $l$  lebih mudah daripada soalan ke  $k$  jika  $\bar{m}_k < \bar{m}_l$ . Seterusnya, pemberat bagi soalan  $j$ ,  $w_j$ , boleh ditentukan dengan menggunakan nilai-nilai min markah  $\bar{m}_j$ :

$$w_j = f_j(\bar{m}_1, \bar{m}_2, \dots, \bar{m}_J) \text{ tertakluk kepada } \sum_j w_j = J. \quad (1)$$

Akhirnya, markah muktamad bagi pelajar  $i$  diberikan oleh,

$$m_i^* = \sum_j w_{j,m_{i,j}}. \quad (3)$$

#### VARIANS MARKAH AWAL SEBAGAI SUKATAN KESUKARAN

Sebagai alternatif kepada penggunaan min markah awal bagi menyukat kesukaran, varians markah awal:

$$v_j = \frac{1}{I} \sum_i (m_{i,j} - \bar{m}_j)^2, \quad (4)$$

juga boleh digunakan. Misalnya, anggapan seperti berikut dilakukan:

*jika  $\bar{m}_k = \bar{m}_l$  dan  $\frac{\bar{m}_k}{v_k} < \frac{\bar{m}_l}{v_l}$ , maka soalan ke  $k$  lebih*

*sukar dari soalan ke  $l$ .*

*jika  $v_k = v_l$  dan  $\frac{\bar{m}_k}{v_k} < \frac{\bar{m}_l}{v_l}$ , maka soalan ke  $k$  lebih*

*sukar dari soalan ke  $l$ .*

Pemberat  $w_j$  boleh ditentukan sebagai fungsi min dan varians:

$$w_j = f_j(m_1, v_1), (m_2, v_2), \dots, (m_J, v_J) \text{ tertakluk kepada } \sum_j w_j = J. \quad (5)$$

JADUAL 1. Markah awal pelajar

<i>i</i>	Markah awal pelajar <i>i</i> bagi soalan <i>j</i> , $m_{ij}$				Jumlah markah	<i>i</i>	Markah awal pelajar <i>i</i> bagi soalan <i>j</i> , $m_{ij}$				Jumlah markah
	<i>j</i> =1	<i>j</i> =2	<i>j</i> =3	<i>j</i> =4			<i>j</i> =1	<i>j</i> =2	<i>j</i> =3	<i>j</i> =4	
1	25	25	25	25	100	13	22	12	0	20	54
2	25	25	22	25	97	14	12	5	13	18	48
3	25	25	17	25	92	15	0	0	23	25	48
4	25	18	20	25	88	16	25	0	0	20	45
5	16	25	18	22	81	17	17	4	14	9	44
6	25	13	18	25	81	18	25	0	18	0	43
7	12	25	15	18	70	19	0	16	11	13	40
8	25	8	12	24	69	20	6	0	10	18	34
9	10	23	19	13	65	21	10	0	0	19	29
10	14	18	20	10	62	22	6	8	0	14	28
11	22	6	8	21	57	23	12	0	6	6	24
12	12	22	0	20	54						

JADUAL 2. Pemberat bagi soalan menggunakan tiga kaedah berbeza

<i>j</i>	1	2	3	4
$s_j$	1.097	0.822	0.854	1.227
$t_j$	0.990	0.530	0.817	1.663
$u_j$	0.935	0.669	0.991	1.405

#### PENGGREDAN BERPANDUKAN PERSENTIL TABURAN

Lazimnya, penggredan dilakukan dengan memadankan setiap gred kepada selang-selang markah tertentu. Misalnya, markah 90 hingga 100 diberi gred A, 80 hingga 89 diberi gred B dan seterusnya. Sebagai alternatif lain, penggredan dilakukan berdasarkan persentil taburan sebenar markah akhir, sebagai contoh, 10% markah tertinggi diberi gred A, 20% berikutnya diberi gred B, dan seterusnya. Kedua-dua kaedah ini sering dipertikai bila markah akhir bertumpu kepada satu nilai (terlalu tinggi atau terlalu rendah), atau dua nilai tertentu (separuh daripada pelajar memperoleh markah tinggi, manakala separuh lagi memperoleh markah yang rendah). Sebagai contoh ekstrim, katalah 80% daripada bilangan pelajar memperoleh markah 90 ke atas, iaitu dengan kaedah penggredan menggunakan selang, 80% akan mendapat gred A, manakala dengan kaedah penggredan menggunakan persentil sebenar, hanya 1.25% daripada pelajar yang memperoleh markah 90 ke atas akan mendapat A.

Sebagai alternatif kepada dua kaedah penggredan di atas, penggredan boleh juga dilakukan dengan menyesuaikan markah-markah muktamad dengan taburan tertentu (misalnya taburan normal piawai), kemudian persentil taburan yang terhasil dari penyesuaian itu digunakan sebagai panduan untuk penentuan gred.

#### MENYUAI MARKAH-MARKAH DENGAN TABURAN NORMAL

Andaikan, bagi seorang pelajar yang dipilih secara rawak dari populasi *I* pelajar dalam satu peperiksaan tertentu, pemboleh ubah rawak bagi markah muktamad, *M*, bertaburan normal dengan min  $\mu$  dan sisihan piawai  $\delta$ . Maka:

$$\mu = \frac{1}{I} \sum_i m_i^* \delta^2 = \frac{1}{I} \sum_i (m_i^* - \mu)^2, \quad (6)$$

dan bertaburan normal piawai. Penentuan gred bagi pelajar *j* dilakukan dengan membandingkan nilai persentil-persentil tertentu dengan nilai markah muktamad terpiawai,  $\frac{m_i^* - \mu}{\delta^2}$ .

#### MENYUAI MARKAH DENGAN TABURAN SELAIN NORMAL

Dengan mengandaikan *M* bertaburan normal, kaedah penggredan berpandukan persentil menjadi mudah kerana persentil bagi taburan normal piawai memang telah tersedia. Sebaliknya, jika *M* diandaikan bertaburan selain taburan normal, penentuan persentil mungkin agak rumitkan.

Sebagai contoh, andaikan  $M$  bertaburan gamma  $(\alpha, \beta)$ . Dengan kaedah momen, penganggar bagi  $\alpha$  dan  $\beta$  boleh ditentukan seperti berikut:

$$\alpha = \frac{a_1^2}{a_2 - a_1^2}, \text{ dan } \beta = \frac{a_1}{a_2 - a_1^2}, \tag{7}$$

yang  $a_k = \frac{1}{I} \sum_i (m_i^*)^k, k = 1, 2$ . Fungsi kebarangkalian melongkok bagi  $M$  diberikan oleh

$$F(m) = \frac{(\beta m)^\alpha}{\Gamma(\alpha)} \cdot e^{-\beta x} \cdot \sum_{n=0}^{\infty} \frac{(\beta m)^n}{\alpha(\alpha+1)\dots(\alpha+n)}, m > 0. \tag{8}$$

Oleh sebab  $M$  tertakluk markah maksimum  $p.J$ , persentil yang akan digunakan bagi penggredan mesti ditentukan dari fungsi kebarangkalian melongkok terpankaskas:

$$F(m | m < p.J) = \frac{F(m)}{F(p.J)}. \tag{9}$$

Dengan bantuan komputer, persamaan (8) dan (9) boleh dihitung dengan agak mudah.

KAJIAN KES

Kajian kes dilakukan kepada markah sebenar bagi satu kursus tertentu, dengan  $I = 23, J = 4$  dan  $p = 25$ . Markah-markah awal bagi kursus berkenaan dipaparkan dalam Jadual 1.

Bagi pemarkahan dengan min sebagai pemberat, rumus  $w_j = s_j = \frac{J \cdot \bar{m}_j}{\sum_j m_j}$  digunakan, manakala bagi pemarkahan

dengan min dan varians sebagai pemberat,

$w_j = t_j = \left( \frac{J \cdot m_j}{v_j} \right) / \sum_j \frac{m_j}{v_j}$  digunakan. Sebagai tambahan,

pemarkahan dengan pemberat  $w_j = u_j = \left( \frac{J}{v_j} \right) \sum_j \frac{1}{v_j}$  juga

dilakukan. Pemberat yang terhasil ditunjukkan dalam Jadual 2, manakala markah-markah muktamad dipaparkan dalam Jadual 3.

Membandingkan ketiga-tiga pemberat  $s_j, t_j,$  dan  $u_j,$  perbezaan antara markah muktamad dengan markah awal didapati paling ketara dengan pemberat  $t_j$ . Jadual 4 memaparkan markah muktamad beberapa orang pelajar yang dikira dengan pemberat  $t_j$  yang agak besar perbezaannya jika dibandingkan dengan markah awal.

JADUAL 3. Markah muktamad pelajar

i	$m_i$	$m_i^s$	$m_{is} - m_i$	$m_i^t$	$m_{it} - m_i$	$m_i^u$	$m_{iu} - m_i$	i	$m_i$	$m_i^s$	$m_{is} - m_i$	$m_i^t$	$m_{it} - m_i$	$m_i^u$	$m_{iu} - m_i$
1	100	100	0	100	0	100	0	13	54	59	5	61	7	57	3
2	97	97	0	98	1	97	0	14	48	50	2	55	7	53	5
3	92	93	1	93	1	92	0	15	48	50	2	60	12	58	10
4	88	90	2	92	4	90	2	16	45	52	7	58	13	51	6
5	81	80	-1	80	-1	80	-1	17	44	45	1	45	1	45	1
6	81	84	3	88	7	85	4	18	43	43	0	39	-4	41	-2
7	70	69	-1	67	-3	68	-2	19	40	38	-2	39	-1	40	0
8	69	74	5	79	10	74	5	20	34	37	3	44	10	41	7
9	65	62	-3	59	-6	62	-3	21	29	34	5	41	12	36	7
10	62	60	-2	56	-6	59	-3	22	28	30	2	33	5	31	3
11	57	62	5	66	9	62	5	23	24	26	2	27	3	26	2
12	54	56	2	57	3	54	0								

JADUAL 4. Perbezaan markah muktamad dan markah awal bagi pemberat  $t_j$

Pelajar ke $i$	6	8	9	10	11	13	14	15	16	20	21
$m_i^*$	88	69	65	62	57	54	48	48	45	34	29
$m_i$	81	79	59	56	66	61	55	60	58	44	41
$m_i^* - m_i$	7	10	-6	-6	9	7	7	12	13	10	12

Bagi penggredan pula, pertimbangkan empat kaedah penggredan berikut ke atas markah muktamad yang terhasil dari pemarkahan berpemberat  $t_j$ :

(i) Penggredan berdasarkan selang markah:

markah	gred	markah	gred	markah	gred
90 ke atas	A	55 ke 69	C	34 ke bawah	E
70 ke 89	B	35 ke 54	D		

(ii) Penggredan berdasarkan persentil sebenar

(iii) Penggredan berdasarkan persentil dengan menganggar taburan normal

(iv) Penggredan berdasarkan persentil dengan menganggar taburan gamma

Bagi kaedah-kaedah ii, iii dan iv, pemberian gred diberikan berdasarkan jadual berikut:

persentil	gred	persentil	gred	persentil	gred
90 ke atas	A	25 ke 54	C	15 ke bawah	E
55 ke 89	B	15 ke 24	D		

Bagi kaedah (iii), keputusan penganggaran dipaparkan dalam Jadual 5. Bagi kaedah (iv), keputusan penganggaran dipaparkan dalam Jadual 6. Keputusan-keputusan penggredan bagi kaedah (i) - (iv) ditunjukkan dalam Jadual 7 dan Jadual 8. Jadual 9 menunjukkan min dan sisihan piawai bagi kaedah penggredan (i)-(iv), dengan memberi mata gred (4,3,2,1,0) kepada gred (A,B,C,D,E).

JADUAL 5. Penggredan berdasarkan persentil dengan menganggar taburan normal

min = 62.48,		sisihan piawai = 21.10			
$n$	90	55	25	15	
persentil ke $n$	89.5	65	48	41	

JADUAL 6. Penggredan berdasarkan persentil dengan menganggar taburan Gamma

$\alpha = 8.76,$		$\beta = 0.14$			
$n$	90	55	25	15	
persentil ke $n$	84.5	56.5	46.5	40.5	

JADUAL 7. Keputusan penggredan menggunakan empat kaedah berbeza

Gred berdasarkan kaedah						Gred berdasarkan kaedah					
i	mi*	(i)	(ii)	(iii)	(iv)	i	mi*	(i)	(ii)	(iii)	(iv)
1	100	A	A	A	A	13	61	C	B	C	B
2	98	A	A	A	A	14	55	C	C	C	C
3	93	A	B	A	A	15	60	C	C	C	B
4	92	A	B	A	A	16	58	C	C	C	B
5	80	B	B	B	B	17	45	D	C	D	D
6	88	B	B	B	A	18	39	D	D	E	E
7	67	C	B	B	B	19	39	D	D	E	E
8	79	B	B	B	B	20	44	D	D	D	D
9	59	C	C	C	B	21	41	D	D	D	D
10	56	C	C	C	C	22	33	E	E	E	E
11	66	C	B	B	B	23	27	E	E	E	E
12	57	C	C	C	B						

JADUAL 8. Taburan jumlah pelajar dengan empat kaedah berbeza

Kaedah	Bilangan (peratus) yang mendapat gred				
	A	B	C	D	E
(i)	4 (17)	3 (13)	9 (39)	5 (22)	2 (9)
(ii)	2 (9)	8 (34)	7 (31)	4 (17)	2 (9)
(iii)	4 (17)	5 (22)	7 (31)	3 (13)	4 (17)
(iv)	5 (22)	9 (39)	2 (9)	3 (13)	4 (17)

JADUAL 9. Min dan sisihan piawai menggunakan empat kaedah berbeza

kaedah	min mata gred	sisih. piawai mata gred	kaedah	min mata gred	sisih. piawai mata gred
(i)	2.0870	1.1763	(iii)	2.0870	1.3159
(ii)	2.1739	1.0896	(iv)	2.3478	1.4021

### PERBINCANGAN

Hasil kajian kes ini menunjukkan perbezaan yang agak ketara antara pemarkahan *a priori* (yang semata-mata bergantung kepada tanggapan pensyarah sewaktu menyediakan soalan) dengan pemarkahan berpemberat (yang menggunakan prestasi pelajar dalam peperiksaan sebagai panduan). Berdasarkan pengalaman penulis berhadapan dengan kes markah di atas pagar, atau peratus pelajar gagal terlalu tinggi, kaedah pemarkahan berpemberat merupakan satu pilihan yang rasional.

Kaedah-kaedah pemarkahan berpemberat yang digunakan dalam kajian kes di atas memberikan pemberat yang lebih kepada soalan yang lebih mudah. Dalam kes peratus pelajar yang memperoleh markah tinggi terlalu ramai, memberikan pemberat yang lebih kepada soalan yang lebih sukar boleh membantu membezakan yang lebih baik di kalangan yang baik itu.

### KESIMPULAN

Dalam penggredan, kaedah (1) merupakan kaedah yang lazim dipakai. Kaedah (2) merupakan kaedah yang agak biasa digunakan di Amerika Syarikat dengan menggunakan hanya 10% terbaik diberi gred A. Berbanding dengan kaedah (3) dan kaedah (4), keputusan menghasilkan persentil yang bernilai lebih rendah. Ini disebabkan oleh bentuk taburan gamma yang terpencong ke kanan. Jika taburan sebenar gred berbentuk dua mod, mengguna model taburan gamma mungkin boleh mengurangkan bilangan pelajar yang memperoleh gred D dan E, berbanding dengan kaedah-kaedah penggredan lain yang telah dibincangkan.

### PENGHARGAAN

Penulis merakamkan terima kasih kepada Universiti Kebangsaan Malaysia atas pembiayaan melalui geran UKM-PTS-011-2010 dan UKM-PTS-2011-035.

### RUJUKAN

- Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis* (2nd. ed.) New York: Springer-Verlag.
- Birenbaum, M. 1997. Assesment preferences and their relationship to learning strategies and orientation. *Higher Education* 33: 71-84.
- Caulkins, J.P., Larkey, P.D. & Wei, J. 1996. Adjusting GPA to reflect course difficulty. Working paper, Heinz School of Public Policy and Management, Carnegie Mellon University. <http://www.heinz.cmu.edu/research/185full.pdf>
- Duru-Bellat, M. 1996. Social inequalities in French secondary schools: from figures to theories. *British Journal of Sociology of Education*, 17 (3): 341-350.
- Hogg, R.V. & Klugman, S.A. 1984. *Loss Distributions* New York: John Wiley.
- Jandaghi, G. 2010. Assessment of validity, reliability and difficulty indices for teacher-built physics exam questions in first year high school. *Educational Research and Review* 5(11): 651-654.
- Jandaghi, G. & Shaterian, F. 2008. Validity, reliability and difficulty indices for instructor-built exam questions. *Journal of Applied Quantitative Method*. 3(2): 151-155.
- Juridah Johari, Jaafar Sahari, Dzuraidah Abd Wahab, Shahrudin Abdullah, Shahrir Abdullah, Mohd Zaidi Omar & Norhamidi Muhamad. 2011. Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia Social and Behavioral Sciences* 18: 71-80.
- Kniveton, B.H. 1996. Student perceptions of assesment methods. *Assesment & Evaluation in Higher Education* 21(3): 229-237.
- Penny, A.J. & Grover, C. 1996. An analysis of student grade expectations and markers consistency. *Assesment & Evaluation in Higher Education* 21(2): 173-183.
- Sarina Sulaiman, Shahida Sulaiman, Mohamad Nor Azwan Mohamad Lazam. 2006. Pengelasan tahap kesukaran soalan menggunakan rangkaian neural. *Proceedings of the 1st International Malaysian Educational Technology Convention*, pp. 357-363.

- Sim, S.-M. & Rasiah, R.I. 2006. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Ann. Acad. Med. Singapore* 35(2).
- Young, M. & Warrington, M. 1996. Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school. *British Journal of Sociology of Education* 17(3): 299-313.
- Zamri Mahamod, Mohd Sarifuddin Mat Salleh & Anita Abdul Rahman. 2008. Taburan kesukaran item soalan peperiksaan penilaian menengah rendah bahasa Melayu berdasarkan model Rasch. *Seminar Pendidikan Kejuteraan dan Alam Bina 2008*: UKM Bangi.
- Zurawski, R.M. 1998. Making the most of exams: procedures for item analysis. *The National Teaching & Learning Forum*, 7(6): 1-4.

Pusat Pengajian Sains Matematik  
Fakulti Sains dan Teknologi  
Universiti Kebangsaan Malaysia  
43600 UKM Bangi  
Selangor, Malaysia

\*Pengarang untuk surat-menyurat; email: hamizun@ukm.my

Diserahkan: 21 September 2011

Diterima: 25 Oktober 2011