

Heri Retnawati

TEORI RESPONS BUTIR DAN PENERAPANNYA

**Untuk Peneliti, Praktisi Pengukuran dan
Pengujian, Mahasiswa Pascasarjana**

Heri Retnawati

TEORI RESPONS BUTIR DAN PENERAPANNYA

**Untuk Peneliti, Praktisi Pengukuran dan
Pengujian, Mahasiswa Pascasarjana**

TEORI RESPONS BUTIR DAN PENERAPANNYA

Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana

Penulis : Heri Retnawati
Sampul : arteholoc numed
Layout : @bay
Cetakan : Pertama, November 2014
ISBN : 978-602-1547-57-4

Diterbitkan

Nuha Medika

Jl. Sadewa No. 1 Sorowajan Baru, Yogyakarta

Telp. 0812 2815 3789

email: nuhamedika@gmail.com - nuhamedika@yahoo.com

facebook: www.facebook.com/nuhamedika

homepage: www.nuhamedika.gu.ma

© 2014, Hak Cipta dilindungi undang-undang,
dilarang keras menterjemahkan, memfotokopi, atau memperbanyak sebagian atau
seluruh isi buku ini tanpa izin tertulis dari penerbit

Undang-Undang Republik Indonesia Nomor 19 Tahun 2002 tentang Hak Cipta.

Sanksi pelanggaran pasal 72:

1. Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksudkan dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp. 1.000.000,00 (satu juta rupiah), atau pidana paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp. 5.000.000.000,00 (lima milyar rupiah).
2. Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta sebagaimana diumumkan dalam ayat (1), dipidana dengan pidana penjara paling lama 5 tahun dan/atau denda paling banyak Rp. 500.000.000,00 (lima ratus juta rupiah).

ISI DI LUAR TANGGUNG JAWAB PENERBIT DAN PERCETAKAN

TEORI RESPONS BUTIR DAN PENERAPANNYA

**Dr. Nur Rosmi, S.Pd., M.Pd., Pengajar dan
Peneliti, Universitas Pasca Sarjana**

PENGANTAR

Segala puji hanya untuk Allah swt yang telah mengkaruniakan rahmat-Nya sehingga buku ini dapat terselesaikan. Buku ini merupakan referensi untuk peneliti, ahli pengukuran dan pengujian, dan juga mahasiswa pascasarjana yang tertarik dan ining mendalami pengukuran khususnya tentang teori respons butir lanjut. Pada buku ini, dilengkapi dengan contoh penerapannya pada penelitian sehingga dapat memberikan gambaran kepada pembaca mengenai penerapan teori respons butir.

Terimakasih kami ucapkan kepada Bapak Prof. Djemari Madapi, Ph.D., Bapak Prof. Dr. Badrun Kartowagiran, Bapak Dr. Samsul Hadi, M.Pd., M.T., Ibu Kana Hidayati, M.Pd. atas kerjasama yang solid dalam beberapa penelitian sehingga menguatkan pemahaman saya pribadi mengenai teori respons butir. Terimakasih pula kepada Bapak Dr. Haryanto atas diskusi dan penjelasannya mengenai logika fuzzy dalam computerized adaptive testing (CAT). Demikian pula kepada Saudara Fauzan Ahmad, Ahmad Madani, dan Fatma Fauzia, yang selalu memotivasi dengan bertanya, "membaca terus boleh, tetapi mana yang telah ditulis untuk dibaca orang lain?".

Semoga buku ini bermanfaat, dan menambah wacana dan referensi untuk teori respons butir di Indonesia. Tidak lupa, saran dan kritik yang membangun tetap diharapkan untuk perbaikan buku ini selanjutnya.

Yogyakarta, 2 September 2014

Heri Retnawati

DAFTAR ISI

PENGANTAR.....	v
DAFTAR ISI	vii
BAB 1 : ASUMSI-ASUMSI TEORI RESPONS BUTIR	1
* Membuktikan Asumsi Teori Respons Butir	4
BAB 2 TEORI RESPONS BUTIR (UNIDIMENSI)	12
* Nilai Fungsi Informasi	18
* Estimasi Paramter Butir	19
* Estimasi Parameter Kemampuan	21
* Menentukan Kecocokan Model	24
BAB 3 TEORI RESPONS BUTIR POLITOMUS.....	32
BAB 4	
TEORI RESPONS BUTIR MULTIDIMENSI.....	45
BAB 5 PENGEMBANGAN BANK SOAL	61
* Pengertian Bank Soal.....	63
* Perlunya Pengembangan Bank Soal.....	63
* Pengembangan Bank Soal	64
* Permasalahan dalam Pengembangan Bank Soal.....	66
* Contoh Pengembangan Bank Soal	67
BAB 6	
MERAKIT PERANGKAT TES DENGAN MEMANFAATKAN NILAI FUGSI	
INFORMASI	78
BAB 7	
PENYETARAAN (EQUATING) DAN CONCORDANCE.....	90

DAFTAR ISI

PENGANTAR.....	v
DAFTAR ISI	vii
BAB 1 : ASUMSI-ASUMSI TEORI RESPONS BUTIR.....	1
* Membuktikan Asumsi Teori Respons Butir	4
BAB 2 TEORI RESPONS BUTIR (UNIDIMENSI)	12
* Nilai Fungsi Informasi	18
* Estimasi Paramter Butir	19
* Estimasi Parameter Kemampuan	21
* Menentukan Kecocokan Model	24
BAB 3 TEORI RESPONS BUTIR POLITOMUS.....	32
BAB 4	
TEORI RESPONS BUTIR MULTIDIMENSI.....	45
BAB 5 PENGEMBANGAN BANK SOAL	61
* Pengertian Bank Soal.....	63
* Perlunya Pengembangan Bank Soal.....	63
* Pengembangan Bank Soal	64
* Permasalahan dalam Pengembangan Bank Soal.....	66
* Contoh Pengembangan Bank Soal	67
BAB 6	
MERAKIT PERANGKAT TES DENGAN MEMANFAATKAN NILAI FUGSI	
INFORMASI	78
BAB 7	
PENYETARAAN (EQUATING) DAN CONCORDANCE.....	90

A. Pengertian dan Asumsi Menghubungkan Tes-Tes	90
B. Jenis Penyetaraan dan Concordance	95
C. Desain Penyetaraan	96
D. Menghubungkan Skor Tes Berdasarkan Teori Tes Klasik.....	98
E. Menghubungkan Skor Tes Berdasarkan Teori Respons Butir untuk Data Dikotomi.....	104
F. Metode Penyetaraan Pada IRT dengan Data Politomi.....	112

BAB 8

TEORI KEBERFUNGSIAN BUTIR DIFERENSIAL.....	125
A. Pengertian Keberfungsian Butir Diferensial	125
B. Metode Pendeteksian DIF pada Teori Respons Butir Unidimensi.....	127
C. Metode Pendeteksian DIF Data Dikotomi Multidimensi	135
D. Metode Pendeteksian DIF data Poltomous Unidimensi GPCM.....	150

BAB 9

MENENTUKAN BATAS KELULUSAN ACUAN KRITERIA

(STANDARD SETTING)	165
✱ Metode Ebel	166
✱ Metode Angoff Tradisional (1971)	169
✱ Metode Extended Angoff (1997).....	170
✱ Metode grup kontras (<i>contrasting group</i>) dari Nedelsky	182
✱ Metode Pemetaan Butir (<i>Item Mapping</i>)	185

BAB 10

TES ADAPTIF TERKOMPUTERISASI	189
1. Bank Butir	191
2. Aturan memulai (<i>starting rule</i>)	191
3. Algoritma pemilihan butir	192
a. Algoritma linear	192
b. Algoritma pohon.....	193
c. Algoritma fuzzy.....	194
4. Estimasi Kemampuan	196
5. Aturan berhenti (<i>stopping rule</i>)	198



BAB 1

ASUMSI-ASUMSI

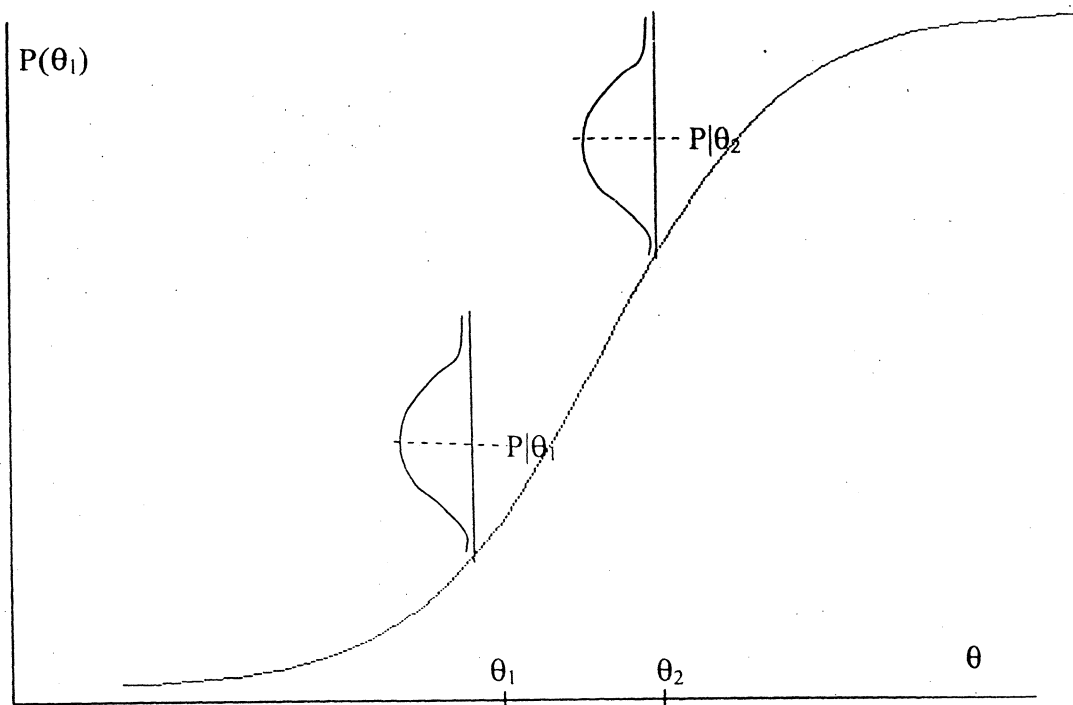
TEORI RESPONS BUTIR



Dalam teori respons butir, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir. Ini berarti bahwa peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta yang mempunyai kemampuan rendah. Hambleton & Swaminathan (1985: 16) dan Hambleton, Swaminathan, & Rogers (1991: 9) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu unidimensi, independensi lokal dan invariansi parameter. Ketiga asumsi dapat dijelaskan sebagai berikut.

Unidimensi, artinya setiap butir tes hanya mengukur satu kemampuan. Contohnya, pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan siswa dalam bidang studi matematika saja, bukan bidang yang lainnya. Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Oleh karena itu, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung satu saja komponen dominan yang mengukur prestasi subjek.

Pada teori respons butir, hubungan antara kemampuan peserta dan skor tes yang dicapai dinyatakan dengan kurva yang tidak linear. Pada Gambar 2 disajikan ilustrasi suatu distribusi kondisional di suatu bagian level kemampuan pada subpopulasi peserta tes. Di sepanjang garis regresi, terdapat sebaran skor tes. Variabilitas kesalahan pengukuran skor tes mungkin terjadi. Jika distribusi bervariasi lintas beberapa subpopulasi, maka tes tidak hanya mengukur kemampuan tunggal saja (Hambleton & Swaminathan, 1985). Misalkan ada 3 subpopulasi, G_1 dan G_2 , skor tes akan disajikan sebagai grafik fungsi yang sama jika tes mengukur satu dimensi kemampuan.



Gambar 1.1

Distribusi Kondisional dari Skor Tes pada Dua Distribusi Kemampuan

Ada beberapa macam cara yang dapat digunakan untuk menguji asumsi unidimensi. Menurut De Mars (2010), ada 3 cara yang sering digunakan, yakni analisis nilai eigen dari matriks korelasi interbutir, uji-Stout pada uji asumsi unidimensi, dan indeks berdasarkan residual pada penyelesaian unidimensi.

Jika faktor-faktor yang mempengaruhi prestasi konstan, maka respons subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain. Kondisi ini disebut dengan independensi lokal. Asumsi independensi lokal ini akan

terpenuhi apabila jawaban peserta terhadap suatu butir soal tidak mempengaruhi jawaban peserta terhadap terhadap butir soal yang lain. Tes untuk memenuhi asumsi independensi lokal dapat dilakukan dengan membuktikan bahwa peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal. Dengan kata lain, korelasi antara pasangan butir terjadi hanya jika kemampuan utama yang diukur dengan sekumpulan butir tidak dipengaruhi oleh suatu kemampuan yang tidak dimodelkan atau kemampuan yang mempengaruhi oleh kedua kumpulan butir tersebut. Menurut De Mars (2010), independensi lokal dapat terdeteksi pula dengan membuktikan asumsi unidimensional.

Menurut Hambleton, Swaminathan, & Rogers (1991: 10), independensi lokal secara matematis dinyatakan sebagai:

$$P(u_1, u_2, \dots, u_n | \theta) = P(u_1 | \theta) \cdot P(u_2 | \theta) \dots \cdot P(u_n | \theta)$$

$$= \prod_{i=1}^n P(u_i | \theta) \dots \dots \dots (1.1)$$

Keterangan :

i : 1, 2, 3, ... n

n : banyaknya butir tes

$P(u_i | \theta)$: probabilitas peserta tes yang memiliki kemampuan dapat menjawab butir ke-i dengan benar.

$P(u_1, u_2, \dots, u_n | \theta)$: probabilitas peserta tes yang memiliki kemampuan dapat menjawab butir ke-1 sampai ke-n dengan benar

Invariansi parameter artinya karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal. Kemampuan seseorang tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesulitannya dan parameter butir tes tidak akan berubah hanya karena diujikan pada kelompok peserta tes yang berbeda tingkat kemampuannya.

Menurut Hambleton, Swaminathan, & Rogers (1991: 18), invariansi parameter kemampuan dapat diselidiki dengan mengajukan dua perangkat tes atau lebih yang

memiliki tingkat kesukaran yang berbeda pada sekelompok peserta tes. Invariansi parameter kemampuan akan terbukti jika hasil estimasi kemampuan peserta tes tidak berbeda walaupun tes yang dikerjakan berbeda tingkat kesulitannya. Invariansi parameter butir dapat diselidiki dengan mengujikan tes pada kelompok peserta yang berbeda. Invariansi parameter butir terbukti jika hasil estimasi parameter butir tidak berbeda walaupun diujikan pada kelompok peserta yang berbeda tingkat kemampuannya.

Dalam teori respons butir, selain asumsi-asumsi yang telah diuraikan sebelumnya, hal penting yang perlu diperhatikan adalah pemilihan model yang tepat. Pemilihan model yang tepat akan mengungkap keadaan yang sesungguhnya dari data tes sebagai hasil pengukuran.

Membuktikan Asumsi Teori Respons Butir

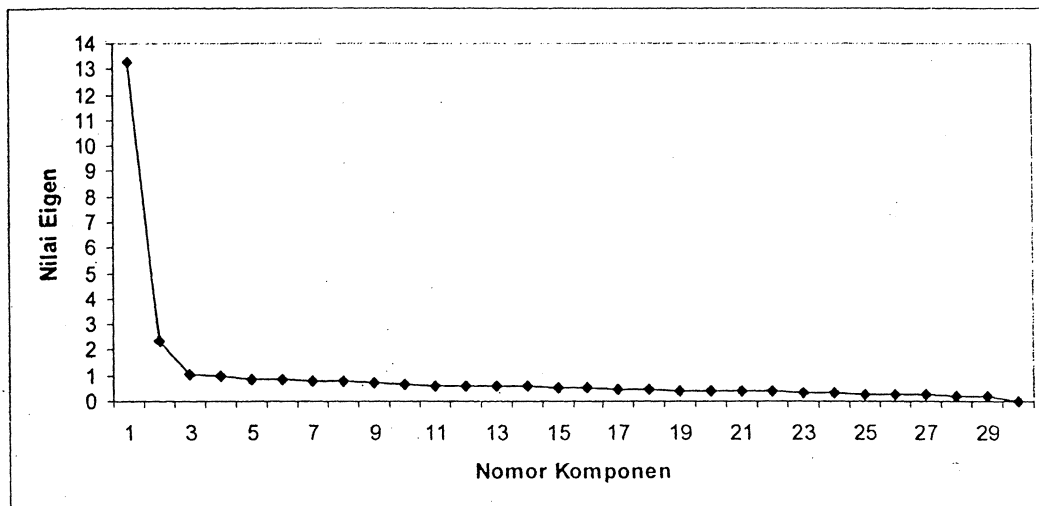
Asumsi unidimensi dapat dibuktikan salah satu diantaranya dengan menggunakan analisis faktor, untuk melihat nilai eigen pada matriks varians kovarians inter-butir. Analisis data dengan analisis faktor didahului dengan analisis kecukupan sampel. Pada tulisan ini akan dibuktikan asumsi unidimensi pada data respons peserta tes terhadap ujian nasional (UN) mata pelajaran matematika tahun 2006 (Heri Retnawati, 2008).

Berdasarkan analisis tentang kecukupan sampel menunjukkan nilai Khi-kuadrat pada uji Bartlett sebesar 21863,839 dengan derajat kebebasan 435 dan nilai-p kurang dari 0,01. Hasil ini menunjukkan bahwa ukuran sampel sebesar 3.012 yang digunakan pada penelitian ini telah cukup.

Tabel 1.1
Hasil Uji KMO dan Bartlett

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.962
Bartlett's Test of Sphericity	Approx. Chi-Square	21863.839
	df	435
	Sig.	.000

Berdasarkan hasil analisis faktor dengan menggunakan SAS/IML, dapat diperoleh bahwa data respons siswa terhadap tes UN Matematika SMP memuat 4 nilai Eigen yang lebih besar dari 1, sehingga dapat dikatakan bahwa tes UN Matematika SMP memuat 4 faktor. Dari keempat faktor ini, ada 59,14% varians yang dapat dijelaskan. Selanjutnya signifikansi dari faktor-faktor ini diuji dengan menggunakan uji χ^2 .



Gambar 1.2

Scree Plot Hasil Analisis Faktor Eksploratori

Nilai Eigen selanjutnya dapat disajikan dengan *scree plot* pada Gambar 1.2. Mencermati hasil *scree plot* tersebut, nampak nilai Eigen mulai landai pada faktor ke-3. Ini menunjukkan bahwa terdapat 1 faktor dominan pada perangkat tes matematika, 1 faktor lainnya juga memberikan sumbangan yang cukup besar terhadap komponen varians yang dapat dijelaskan (banyaknya lereng yang curam pada *scree plot* menunjukkan banyaknya dimensi). Mulai faktor ketiga dan seterusnya, pada grafik menunjukkan sudah mulai mendatar. Hal ini menunjukkan bahwa perangkat tes matematika mengukur paling tidak 2 faktor dengan faktor yang pertama merupakan faktor dominan.

Tabel 1.2
 Nilai Eigen dan Komponen Varians Hasil Analisis Faktor

No. Komponen	Nilai Eigen	Perbedaan Nilai Eigen	Proporsi	Kumulatif
1	13,2860	10,9153	0,4429	0,4429
2	2,3707	1,2986	0,0790	0,5219
3	1,0721	0,0587	0,0357	0,5576
4	1,0134	0,1385	0,0338	0,5914
5	0,8749	0,0512	0,0292	0,6206
6	0,8237	0,0467	0,0275	0,6480
7	0,7770	0,0216	0,0259	0,6739
8	0,7554	0,0400	0,0252	0,6991
9	0,7154	0,0422	0,0238	0,7230
10	0,6732	0,0607	0,0224	0,7454
11	0,6126	0,0116	0,0204	0,7658
12	0,6010	0,0191	0,0200	0,7859
13	0,5819	0,0232	0,0194	0,8052
14	0,5587	0,0285	0,0186	0,8239
15	0,5303	0,0269	0,0177	0,8415
16	0,5034	0,0220	0,0168	0,8583
17	0,4814	0,0430	0,0160	0,8744
18	0,4384	0,0261	0,0146	0,8890
19	0,4124	0,0206	0,0137	0,9027
20	0,3917	0,0233	0,0131	0,9158
21	0,3685	0,0067	0,0123	0,9281
22	0,3618	0,0219	0,0121	0,9401
23	0,3399	0,0141	0,0113	0,9515
24	0,3258	0,0443	0,0109	0,9623
25	0,2815	0,0253	0,0094	0,9717
26	0,2561	0,0032	0,0085	0,9802
27	0,2529	0,0580	0,0084	0,9887
28	0,1949	0,0174	0,0065	0,9952
29	0,1774	0,2098	0,0059	1,0011
30	-0,0323		-0,0011	1,0000

Cara lain yang dapat dilakukan untuk mengetahui banyaknya faktor yang termuat yakni dengan membandingkan nilai khi-kuadrat dari tiap faktor pada analisis faktor. Nilai khi-kuadrat pada analisis ini dilakukan dengan bantuan program TESTFACT. Dengan melakukan analisis faktor dengan memasukkan 1 faktor saja, diperoleh nilai Khi-kuadrat sebesar 33353,97 dan derajat kebebasan sebesar 2951,00. Pada 2 faktor, diperoleh nilai Khi-kuadrat sebesar 33124,35 dan derajat kebebasan 2922,00 dan dengan 3 faktor sebesar 33006,25 dengan derajat kebebasan 2894,00. Terakhir, jika memasukkan 4

faktor diperoleh nilai Khi-kuadrat sebesar 36387,73 dengan derajat kebebasan 2867,00. Selanjutnya dapat dihitung selisih nilai khi-kuadrat untuk mengetahui model mana yang lebih baik. Hasil pengujian selengkapnya disajikan pada Tabel 1.3.

Berdasarkan hasil analisis ini, dapat disimpulkan bahwa berdasarkan data empiris, perangkat tes UN mengukur dengan lebih baik berturut-turut 3 faktor, 2 faktor, dan 1 faktor. Faktor-faktor inilah yang selanjutnya disebut sebagai dimensi. Hasil berdasarkan uji statistik ini memperkuat hasil penentuan faktor dengan menggunakan *scree-plot*, yang menunjukkan bahwa tes UN Matematika SMP mengukur 2 dimensi, namun berdasarkan statistik, ada 3 dimensi yang menjadi variabel yakni 1,2, dan 3 dimensi. Hal ini menunjukkan bahwa perangkat tes tidak unidimensi, namun multidimensi.

Tabel 1.3
Hasil Uji Khi-Kuadrat untuk Menentukan Banyaknya Muatan Faktor

Banyaknya Faktor	χ^2	df	$\chi^2_{(k)} - \chi^2_{(k+1)}$	$df_{(k)} - df_{(k+1)}$	$\chi^2_{\text{kritis } (0,05, df)}$	Kesimpulan
1	33353,97	2951				
2	33124,35	2922	229,62	29	42,56	Model 2 faktor lebih baik dibanding model 1 faktor
3	33006,25	2894	118,10	28	41,34	Model 3 faktor lebih baik dibanding model 2 faktor
4	36387,73	2867	-3381,48	27	40,11	Model 4 faktor tidak lebih baik dibanding model 3 faktor

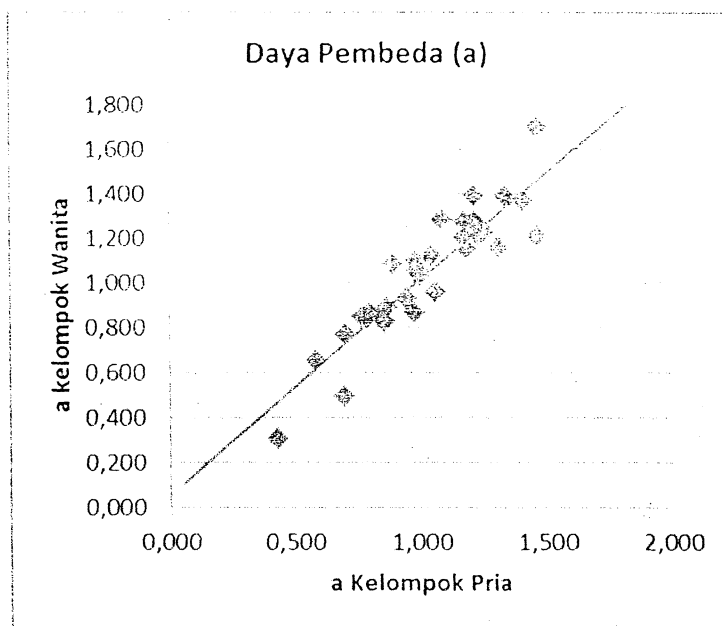
Asumsi kedua merupakan asumsi independensi lokal. Asumsi ini otomatis terbukti, setelah dibuktikan dengan unidimensionalitas data respons peserta terhadap suatu tes. Jika terbukti hasil analisis multidimensi (tidak unidimensi), antar faktor juga belum tentu ada korelasi. Hambleton pada tahun 1991 mengembangkan prosedur menguji asumsi independensi local dari dua butir dengan khi-kuadrat. Namun untuk butir yang cukup banyak, cara ini menjadi kurang efisien.

Asumsi yang ketiga adalah invariansi parameter butir dan parameter kemampuan. Asumsi ini dibuktikan dengan mengestimasi parameter butir pada kelompok peserta tes yang berbeda, misalnya kelompok berdasarkan jenis kelamin, tempat tinggal, status social ekonomi, dan lain-lain. Dari hasil estimasi, baik parameter daya pembeda (a), tingkat kesulitan (b), dan pseudo guessing (c) pada kedua kelompok kemudian disajikan

dalam diagram pencar. Jika berkorelasi tinggi, atau titik-titik pada diagram pencar mendekati garis yang melewati titik asal dengan gradien atau kemiringan 1, maka dianggap parameter-parameter tersebut invarian.

Sebagai contoh suatu soal ujian matematika yang terdiri dari 30 butir, peserta dikelompokkan berdasarkan jenis kelaminnya, kelompok laki-laki dan kelompok perempuan. Parameter a , b , dan c masing-masing kelompok diestimasi. Untuk parameter a kelompok laki-laki dan kelompok perempuan digambarkan dalam sebuah diagram pencar (*scatter plot*), demikian pula halnya dengan parameter b dan parameter c .

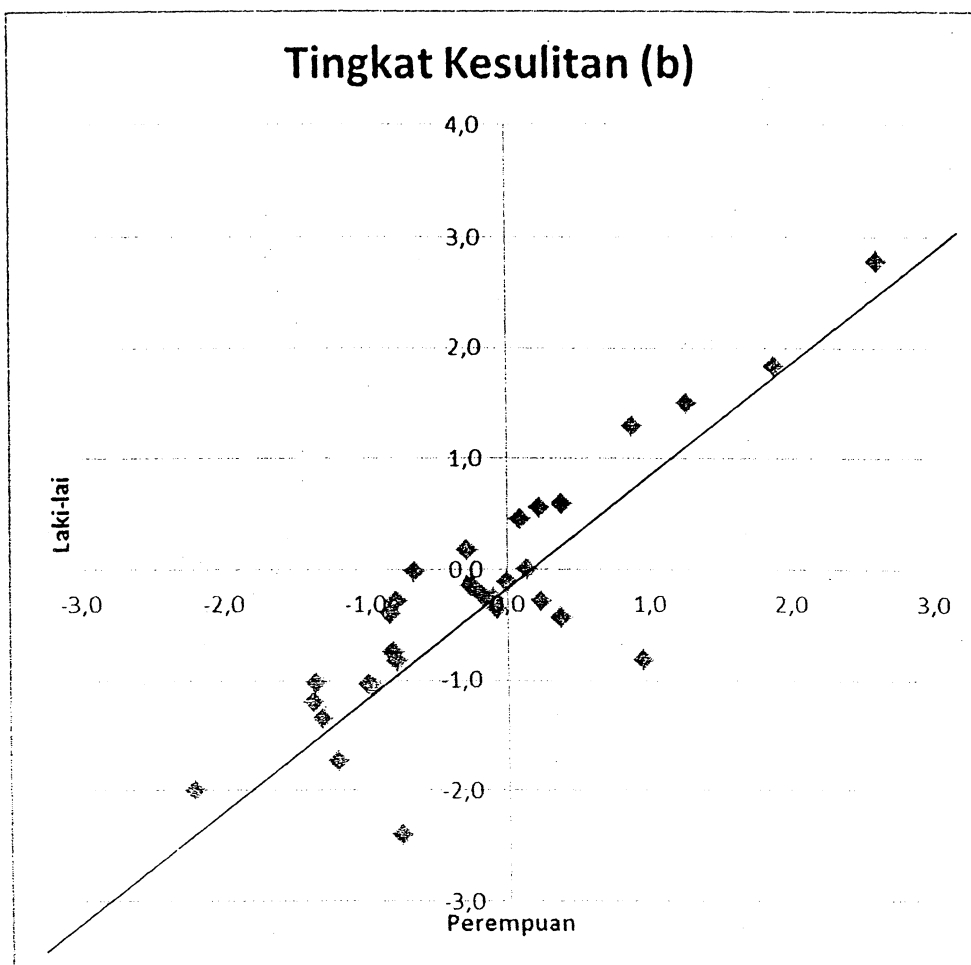
Diagram pencar untuk daya pembeda, yang diestimasi pada kelompok wanita dan kelompok pria digambarkan pada Gambar 1.3. Berdasarkan gambar tersebut, diperoleh bahwa masing-masing titik berada relative dekat dengan garis dengan kemiringan 1. Hal ini menunjukkan bahwa tidak terjadi variasi parameter hasil estimasi pada kelompok wanita dan kelompok pria.



Gambar 1.3
Invariansi Parameter a kelompok pria dan kelompok wanita

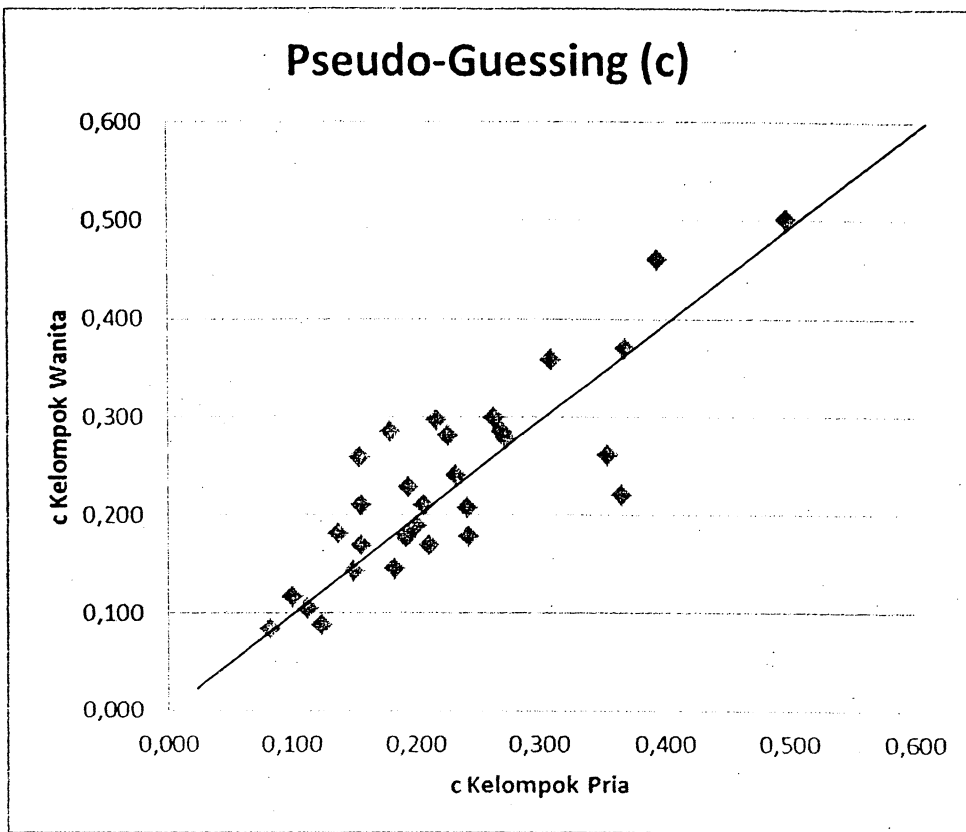
Demikian pula pada diagram pencar untuk parameter tingkat kesulitan dan parameter tebakan semu, yang diestimasi pada kelompok wanita dan kelompok pria digambarkan pada Gambar 1.4 dan 1.5. Berdasarkan gambar tersebut, diperoleh bahwa masing-masing titik berada relative dekat dengan garis dengan kemiringan 1. Hal ini menunjukkan bahwa

tidak terjadi variasi parameter b dan c hasil estimasi pada kelompok wanita dan kelompok pria. Dengan kata lain, invariansi parameter terpenuhi.

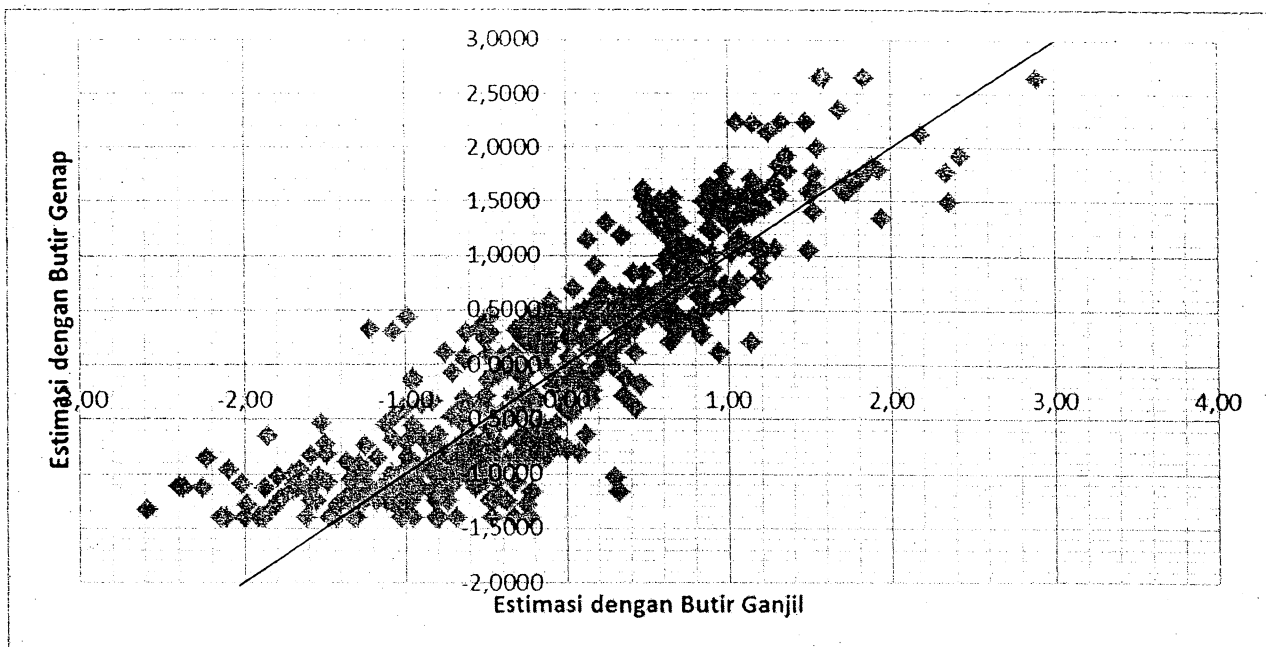


Gambar 1.4
Invariansi Parameter b kelompok pria dan kelompok wanita

Dengan cara yang sama, invariansi parameter kemampuan dapat dibuktikan dengan terlebih dahulu mengestimasi parameter kemampuan menggunakan butir nomor genap dan estimasi parameter kemampuan menggunakan parameter butir yang ganjil. Klasifikasi dapat pula dilakukan menggunakan separuh butir bagian atas dan separuh butir bagian bawah. Selanjutnya dibuat diagram pencar, kemudian dibandingkan kedekatannya dengan garis dengan kemiringan 1. Pada buku ini disajikan diagram pencar 500 peserta tes dengan menggunakan klasifikasi butir ganjil dan butir genap. Hasilnya disajikan pada Gambar 1.6. Berdasarkan hasil analisis invariansi parameter a, b, c dan σ^2 , maka diperoleh hasil bahwa invariansi parameter terbukti.



Gambar 1.5
Invariansi Parameter c kelompok pria dan kelompok wanita



Gambar 1.6
Invariansi Parameter Kemampuan Mengerjakan Butir-butir Ganjil dan Genap

Daftar Pustaka

DeMars, C.E. (2010). *Item response theory*. New York: Oxford University Press.

Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: SSI.

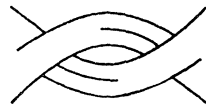
Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.

* * *

BAB 2

TEORI RESPONS BUTIR (UNIDIMENSI)



Pada pengukuran dalam pendidikan, kesehatan, psikologi, dan lainnya, penskoran sering dilakukan secara dikotomi. Misalnya dalam evaluasi yang dilaksanakan dalam pendidikan, siswa menjawab butir soal suatu tes yang berbentuk pilihan ganda dengan benar, biasanya diberi skor 1 dan 0 jika menjawab salah. Pada penyekoran dengan pendekatan teori tes klasik, kemampuan siswa dinyatakan dengan skor total yang diperolehnya. Prosedur ini kurang memperhatikan interaksi antara setiap orang siswa dengan butir.

Pendekatan teori respons butir merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Ada dua prinsip yang digunakan pada pendekatan ini, yakni prinsip relativitas dan prinsip probabilitas. Pada prinsip relativitas, unit dasar dari pengukuran bukanlah siswa atau butir, tetapi lebih kepada performance siswa relatif terhadap butir. Jika β_n merupakan indeks dari kemampuan siswa ke n pada trait yang diukur, dan δ_i merupakan indeks dari tingkat kesulitan dari butir ke- i relatif yang terkait dengan kemampuan yang diukur, maka bukan β_n atau δ_i yang merupakan unit pengukuran, tetapi lebih kepada perbedaan antara kemampuan dan dari siswa relative

terhadap tingkat kesulitan butir atau $(\beta_n - \delta_i)$ perlu dipertimbangkan. Sebagai alternatifnya perbandingan antara kemampuan terhadap tingkat kesulitan dapat digunakan. Jika kemampuan dari siswa melampaui tingkat kesulitan butir, maka respons siswa diharapkan benar, dan jika kemampuan siswa kurang dari tingkat kesulitan butir, maka respons siswa diharapkan salah (Keeves dan Alagumalai, 1999:24).

Pada teori respons butir, prinsip probabilitas menjadi perhatian. Misalkan kemampuan siswa ke n dinyatakan dengan θ_n dan tingkat kesulitan dari butir dinyatakan dengan Δ_i maka sesuai dengan prinsip relativitas, jika $\theta_n > \Delta_i$ siswa diharapkan menjawab dengan benar, dan $\theta_n < \Delta_i$ siswa diharapkan menjawab salah. Lebih jauh lagi, jika kemungkinan (odds) dari respons siswa terhadap butir diberikan oleh $\frac{\theta_n}{\Delta_i} > 1$ (siswa diharapkan menjawab dengan benar), $\frac{\theta_n}{\Delta_i} < 1$, siswa diharapkan menjawab salah, dan

$\frac{\theta_n}{\Delta_i} = 1$, akan terjadi jika kesempatan 50% menjawab benar.

Jika p_{ni} merupakan peluang menjawab benar, maka $1 - p_{ni}$ merupakan respons tidak benar dan odds untuk respons diberikan oleh

$$\frac{\theta_n}{\Delta_i} = \frac{p_{ni}}{1 - p_{ni}} \text{ dan } \frac{\theta_n}{\Delta_i} = 1 \text{ terjadi jika } p_{ni} = 0,5 \dots \dots \dots (2.1)$$

Probabilitas respons menjawab benar berada pada rentang 0 sampai dengan 1.0 dan hal ini menghalangi data dinyatakan sebagai skala interval. Skor mentah yang dihasilkan dari cara ini sulit dinyatakan sebagai skala. Untuk mengatasi permasalahan ini, dapat digunakan transformasi logistik, yang melibatkan logaritma natural dari odds.

$$\ln \frac{\theta_n}{\Delta_i} = \ln \left(\frac{p_{ni}}{1 - p_{ni}} \right) \dots \dots \dots (2.2)$$

yang senilai dengan

$$\ln \theta_n - \ln \Delta_i = \ln \left(\frac{p_{ni}}{1 - p_{ni}} \right) \dots \dots \dots (2.3)$$

Misalkan $\ln \theta_n = \beta_n$ dan $\ln \Delta_i = \delta_i$ maka

$$\beta_n - \delta_i = \ln \left(\frac{p_{ni}}{1 - p_{ni}} \right) \dots \dots \dots (2.4)$$

Persamaan tersebut senilai dengan

$$\frac{p_{ni}}{1 - p_{ni}} = \exp(\beta_n - \delta_i) \dots \dots \dots (2.5)$$

atau dengan kata lain probabilitas respons menjawab benar ($x_{ni} = 1$) dapat ditulis sebagai

$$p_{ni} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \dots \dots \dots (2.6)$$

dan probabilitas respons menjawab salah

$$p_{ni}(x_{ni} = 0) = 1 - p_{ni}(x_{ni} = 1) = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \dots \dots \dots (2.7)$$

Model ini merupakan model logistik univariat (Hosmer dan Lemeshow, 1989).

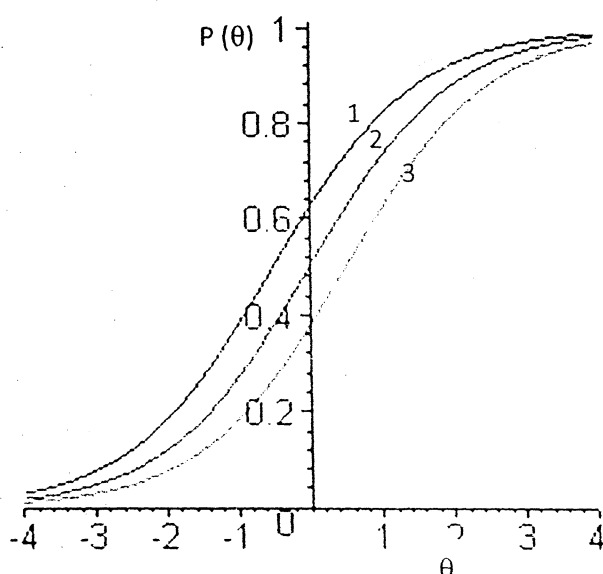
Bentuk persamaan yang lebih dikenal dalam pengukuran untuk model ini, yang biasa disebut dengan model Rasch (Hambleton, Swaminathan, dan Rogers, 1991: 12) sebagai berikut :

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \text{ , dengan } i : 1, 2, 3, \dots, n \dots \dots \dots (2.8)$$

- $P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir i dengan benar
- θ : tingkat kemampuan subyek (sebagai variabel bebas)
- b_i : indeks kesukaran butir ke- i
- e : bilangan natural yang nilainya mendekati 2,718
- n : banyaknya butir dalam tes

Parameter b_i merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Misalkan suatu butir tes mempunyai parameter $b_i = 0,3$, artinya diperlukan kemampuan minimal 0,3 pada skala untuk dapat menjawab benar dengan peluang 50%. Semakin besar nilai parameter b_i , maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b_i , maka makin sulit butir soal tersebut.

Hubungan peluang menjawab benar $P_i(\theta)$ dengan tingkat kemampuan peserta (θ) dapat digambarkan sebagai kurva karakteristik butir (*item characteristic curve, ICC*). Gambar 1 berikut merupakan ilustrasi kurva karakteristik butir untuk model Rasch (1 parameter, 1P), dengan butir 1 ($b=-0,5$), butir 2 ($b=0$) dan butir 3 ($b=0,5$).



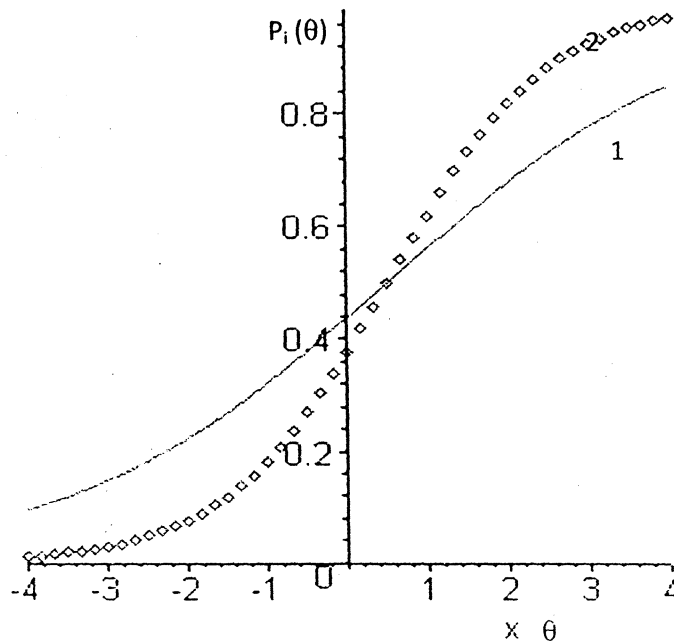
Gambar 2.1. Kurva karakteristik butir untuk model 1P, dengan butir 1 ($b=-0,5$), butir 2 ($b=0$) dan butir 3 ($b=0,5$)

Jika $(\theta - b_i)$ ditransformasi menjadi $a_i(\theta - b_i)$ dengan a_i suatu konstanta, maka a_i ini merupakan tingkat daya pembeda butir (*item difficulty*). Selanjutnya model ini merupakan model logistik 2 parameter (2P) dengan parameter butir yaitu indeks kesukaran butir (b_i) dan indeks daya beda butir (a_i), yang memenuhi :

$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \text{ , dengan } i : 1, 2, 3, \dots, n \dots \dots \dots (2.9)$$

Jika a_i bernilai 1, maka model 2 parameter ini menjadi model logistik 1 parameter.

Sebagai ilustrasi, kurva karakteristik butir 1 ($a=0,5$; $b=0,5$) dan butir 2 ($a=1$; $b=0,5$) disajikan pada Gambar 2.2.

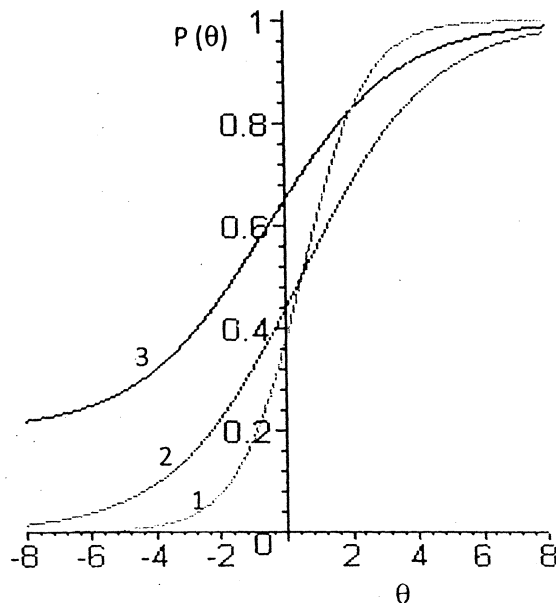


Gambar 2.2. Kurva karakteristik butir model 2P, dengan butir 1 ($a=0,5$; $b=0,5$) dan butir 2 ($a=1$; $b=0,5$)

Pada gambar 2 tersebut, asimtot kiri (untuk $\theta \rightarrow -\infty$) adalah 0. Jika bukan 0, maka nilai ini merupakan parameter tebakan semu (*pseudo guessing*) atau c_i , sehingga model logistik menjadi model 3 parameter (3P). Dengan adanya tebakan semu pada model logistik tiga parameter, memungkinkan siswa yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Secara matematis, model logistik tiga parameter dapat dinyatakan sebagai berikut (Hambleton, Swaminathan, dan Rogers, 1991: 17, Hambleton, dan Swaminathan, 1985 : 49, Van der Linden dan Hambleton, 1997: 13).

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \dots\dots\dots (2.10)$$

Sebagai ilustrasi, gambar 2.3 merupakan kurva karakteristik butir 1 ($a=1, b=0,5, c=0$), butir 2 ($a=0,5, b=0,5, c=0$) dan butir 3 ($a=0,5, b=0,5, c=0,2$).



Gambar 2.3. kurva karakteristik butir model 3P, dengan butir 1 ($a=1, b=0,5, c=0$), butir 2 ($a=0,5, b=0,5, c=0$) dan butir 3 ($a=0,5, b=0,5, c=0,2$)

Nilai kemampuan peserta (θ) terletak di antara -4 dan $+4$, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Secara teoretis, nilai b_i terletak di antara -2 dan $+2$. Suatu butir dikatakan baik jika nilai ini berkisar antara -2 dan $+2$ (Hambleton dan Swaminathan, 1985: 107). Jika nilai b_i mendekati -2 , maka indeks kesukaran butir sangat rendah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesukaran butir sangat tinggi untuk suatu kelompok peserta tes.

Parameter a_i merupakan daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik, a_i merupakan kemiringan (*slope*) dari kurva di titik b_i pada skala kemampuan tertentu. Karena merupakan kemiringan, diperoleh semakin besar kemiringannya, maka semakin besar daya pembeda butir tersebut. Secara teoretis, nilai a_i ini terletak antara -2 dan $+2$. Pada pada butir yang baik nilai ini mempunyai hubungan positif dengan

performen pada butir dengan kemampuan yang diukur, dan a_i terletak antara 0 dan 2 (Hambleton dan Swaminathan, 1985: 37).

Peluang menjawab benar dengan memberikan jawaban tebakan semu dilambangkan dengan c_i , yang disebut dengan *tebakan semu*. Parameter ini memberikan suatu kemungkinan asimtot bawah yang tidak nol (*nonzero lower asymptote*) pada kurva karakteristik butir (ICC). Parameter ini menggambarkan probabilitas peserta dengan kemampuan rendah menjawab dengan benar pada suatu butir yang mempunyai indeks kesukaran yang tidak sesuai dengan kemampuan peserta tersebut. Besarnya harga c_i diasumsikan lebih kecil daripada nilai yang akan dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir. Pada suatu butir tes, nilai c_i ini berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai c_i tidak lebih dari $1/k$, dengan k banyaknya pilihan (Hullin, 1983: 36).

Nilai Fungsi Informasi

Dalam teori respons butir, dikenal nilai fungsi informasi. Fungsi informasi butir (*Item Information Functions*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes, dan perbandingan beberapa perangkat tes. Fungsi informasi butir menyatakan kekuatan atau sumbangan butir tes dalam mengungkap latent trait yang diukur dengan tes tersebut. Dengan fungsi informasi butir diketahui butir yang mana yang cocok dengan model sehingga membantu dalam seleksi butir tes. Secara matematis, fungsi informasi butir memenuhi persamaan sebagai berikut.

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \dots\dots\dots (2.11)$$

keterangan :

i : 1,2,3,...,n

$I_i(\theta)$: fungsi informasi butir ke-i

$P_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

$P'_i(\theta)$: turunan fungsi $P_i(\theta)$ terhadap θ

$Q_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton dan Swaminathan, 1985: 94). Berhubungan dengan hal ini, fungsi informasi perangkat tes akan tinggi jika butir tes mempunyai fungsi informasi yang tinggi pula. Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

$$I_i(\theta) = \sum_{i=1}^n I_i(\theta) \dots\dots\dots (2.12)$$

Nilai-nilai indeks parameter butir dan kemampuan peserta merupakan hasil estimasi. Karena merupakan hasil estimasi, maka kebenarannya bersifat probabilitas dan tidak terlepas dengan kesalahan pengukuran. Dalam teori respon butir, kesalahan penaksiran standar (*Standard Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil atau sebaliknya (Hambleton, Swaminathan dan Rogers, 1991, 94). Jika nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ dan nilai estimasi *SEM* dinyatakan dengan $SEM(\hat{\theta})$, maka hubungan keduanya, menurut Hambleton, Swaminathan, dan Rogers (1991 : 94) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (2.13)$$

Estimasi Paramter Butir

Pada model 1P, 2P dan 3P, untuk menganalisis jawaban siswa, yang perlu menjadi perhatian adalah pengestimasiian parameter butir dan parameter kemampuan peserta. Dalam pengestimasiian ini, dikenal fungsi likelihood. Fungsi likelihood untuk kasus dengan N siswa dan n butir dapat dinyatakan dengan

$$L(\theta, b;u) = \prod_i \prod_j P_i(\theta_j; b_i)^{u_{ij}} [1 - P_i(\theta_j; b_i)]^{l-u_{ij}} \quad (1P)$$

$$L(\theta, a,b;u) = \prod_i \prod_j P_i(\theta_j; a_i, b_i)^{u_{ij}} [1 - P_i(\theta_j; a_i, b_i)]^{l-u_{ij}} \quad (2P) \dots\dots\dots (2.14)$$

$$L(\theta, a,b,c;u) = \prod_i \prod_j P_i(\theta_j; a_i, b_i, c_i)^{u_{ij}} [1 - P_i(\theta_j; a_i, b_i, c_i)]^{l-u_{ij}} \quad (3P)$$

Selanjutnya diestimasi nilai-nilai yang memaksimumkan fungsi ini. Prosedur yang dapat dipilih yakni prosedur likelihood maksimum gabungan (*joint maximum likelihood, JML*) atau prosedur likelihood maksimum marginal (*marginal maximum likelihood, MML*) atau juga dengan pendekatan Bayes.

Untuk mengestimasi parameter-parameter butir pada model logistik 1P, 2P maupun 3P, ada beberapa perangkat lunak yang dapat digunakan, diantaranya Rascal (1P), Ascal (2P dan 3P), Bilog, Xalibrate dan Multilog. Keluaran (*output*) dari program-program ini juga menyediakan hasil pengestimasian parameter peserta tes.

Langkah selanjutnya adalah mengetahui kecocokan model dari data yang dianalisis. Uji statistik untuk kecocokan model salah satunya uji perbandingan likelihood (*likelihood ratio test*). Uji ini digunakan untuk mengecek apakah estimasi parameter butir dalam grup skor yang berbeda bernilai sama pada kesalahan penyampelan dari estimasi. Secara teoritis, responden yang berukuran N dapat dibuat menjadi interval-interval pada skala kontinum untuk θ , yang merupakan dasar untuk mengestimasi nilai θ . Statistik Khi-kuadrat dari perbandingan *likelihood* digunakan untuk membandingkan frekuensi menjawab benar dan tidak benar dari respons pada interval yang diharapkan dari model yang cocok pada rata-rata interval $\hat{\theta}_h$, dengan persamaan :

$$G_j^2 = 2 \sum_{h=1}^{n_g} \left[r_{hj} \ln \frac{r_{hj}}{N_h P_j(\bar{\theta}_h)} + (N_h - r_{hj}) \ln \frac{N_h - r_{hj}}{N_h [1 - P(\bar{\theta}_h)]} \right] \dots\dots\dots(2.15)$$

dengan n_g merupakan banyaknya interval, r_{hj} merupakan frekuensi respons yang benar untuk butir pada interval h , N_h merupakan banyaknya anggota sampel yang berada dalam interval, dan $P_j(\bar{\theta}_h)$ merupakan nilai dari fungsi respons sesuai model untuk butir j pada $\bar{\theta}_h$, yang merupakan kemampuan rata-rata responden pada interval h (Mislevy dan Bock, 1990).

Pada program Bilog, untuk menentukan banyaknya interval yang dibuat pada skala kontinu untuk θ , mula-mula dibuat maksimum 20 interval. Setiap responden disarangkan pada interval tersebut termasuk estimasi EAP (*expected a posteriori*), berdasarkan tipe prior yang dispesifikasikan pemakai dari skor yang diperoleh responden. Pada setiap butir tes, probabilitas harapan dari respons yang sesuai dengan estimasi rata-rata EAP untuk

kemampuan dari kasus yang berada dalam interval digunakan sebagai proporsi harapan untuk interval tersebut.

Khi-kuadrat perbandingan kemungkinan dihitung setelah mengkombinasikan interval-interval yang ekstrim, hingga frekuensi harapan pada gabungan interval-interval tersebut lebih dari lima. Derajat kebebasan dari khi-kuadrat perbandingan kemungkinan sama dengan banyaknya interval-interval yang telah dikombinasikan (Mislevy dan Bock, 1990).

Estimasi Parameter Kemampuan

Agar informasi yang diperoleh berguna dalam penskoran tes, parameter butir perlu diestimasi. Estimasi parameter butir dan mengecek kecocokan model sering disebut sebagai kaliberasi butir. Kaliberasi ini dapat dilakukan jika data respons peserta terhadap tes telah diperoleh. Paling tidak ada 2 pendekatan yang dapat digunakan untuk estimasi parameter butir atau melakukan kaliberasi butir, yakni estimasi Marginal Maximum Likelihood (MML) dan estimasi Marginal Maximum A Posteriori (MMAP) (Du Toit, 2006).

MML merupakan metode yang diyakini efisien untuk semua model respons butir dan untuk tes yang panjang maupun yang pendek. MML mengasumsikan adanya respons yang berbeda dari kemampuan θ yang sama. Asumsi ini memungkinkan untuk menghitung probabilitas dari sebagian pola skor butir

$$X = (x_1, x_2, x_3, \dots, x_n)$$

Pada seorang peserta tes dengan kemampuan θ , probabilitas dinyatakan dengan

$$P(x|\theta) = \prod_{j=1}^n [P_j(\theta)]^{x_j} [1 - P_j(\theta)]^{1-x_j} \dots\dots\dots (2.16)$$

Dengan $P_j(\theta)$ merupakan peluang menjawab benar butir ke-j, dan $1 - P_j(\theta)$ merupakan peluang menjawab salah. $P(x|\theta)$ merupakan probabilitas dari pola jawaban x dengan syarat pada suatu θ . Hal ini membedakannya dengan probabilitas dari suatu pola x dari kemampuan yang belum diketahui dari suatu populasi acak dengan θ diperoleh dari suatu fungsi kerapatan kontinu $g(\theta)$.

Kondisi ini disebut dengan probabilitas tidak bersyarat yang diberikan oleh integral tertentu

$$P(x) = \int_{\sim} P(x|\theta)g(\theta) d\theta$$

yang menunjukkan probabilitas marginal dari x. Karena kemampuan θ telah dikelurakan, maka besaran ini hanya ditentukan oleh parameter butir saja.

Pada penerapan teori respons butir, integral ini tidak dapat dinyatakan dalam bentuk tertutup tetapi probabilitas marginalnya dapat diketahui menggunakan rumus kuadratur dari Gauss

$$\bar{P}_x \approx \sum_{k=1}^q P(x|X_k)A(X_k)$$

Dengan X merupakan titik kuadratur dan $A(X_k)$ merupakan pembobotan positif terkait dengan fungsi kerapatan $g(X)$. Banyaknya titik kuadratur yang direkomendasikan yakni 2 kali lipat dari akar kuadrat banyaknya butir sebagai banyaknya titik kuadratur maksimum.

Dalam metode MML, nilai parameter butir dipilih yang dapat memaksimumkan logaritma dari fungsi marginal maximum likelihood yang didefinisikan sebagai

$$\log L_m = \sum_{i=1}^S r_i \log_e \bar{P}(x_i) \dots \dots \dots (2.17)$$

Dengan r_i merupakan frekuensi dari pola x_i yang diamati dari ukuran sampel sebesar N peserta dan S merupakan banyaknya pola yang berbeda. Pada model 3 parameter, kondisi maksimum diberikan dalam persamaan likelihood

$$\sum_{k=1}^q \left(\frac{\bar{r}_{jk} - \bar{N}_{jk} P_j(X_k)}{P_j(X_k)[1 - P_j(X_k)]} \right) \frac{\partial P_j(X_k)}{\partial \begin{pmatrix} c_j \\ a_j \\ g_j \end{pmatrix}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \dots \dots \dots (2.18)$$

Dengan

$$\bar{r}_{jk} = \sum_l^S r_l x_{lj} P(x_l|X_k) A(X_k) / \bar{P}_{x_l}$$

$$\bar{N}_k = \sum_l^S r_l P(x_l|X_k) A(X_k) / \bar{P}_{x_l}$$

yang berturut-turut merupakan ekspektasi posterior dari banyaknya jawaban benar dan banyaknya usaha menjawab benar di titik X_k dan x_{lj} skor 0-1 butir ke-j pada pola ke-l.

Langkah untuk mengestimasi tersebut disebut algoritma E dan lanjutannya disebut algoritma M, sehingga keseluruhannya disebut dengan EM algoritma.

Pada tes klasik, kemampuan peserta tes diestimasi berdasarkan kemampuan menjawab benar. Pada tRB, kemampuan diestimasi dengan fungsi non linear yang

kemudian disebut dengan skor. Ada 3 metode estimasi yang sering digunakan, yakni estimasi maksimum likelihood, estimasi Bayes, dan estimasi Modal Bayes.

Likelihood maksimum dari skor kemampuan peserta diestimasi dengan memaksimalkan fungsi

$$\log L_i(\theta) = \sum_{j=1}^n \{x_{ij} \log_e P_j(\theta) + (1 - x_{ij}) \log_e [1 - P_j(\theta)]\} \dots \dots \dots (2.19)$$

Dengan fungsi yang cocok dengan butir j.

Selanjutnya diselesaikan persamaan implisit likelihood

$$\frac{\partial \log L_i(\theta)}{\partial \theta} = \sum_{j=1}^n \frac{x_{ij} - P_j(\theta)}{P_j(\theta)[1 - P_j(\theta)]} \cdot \frac{\partial P_j(\theta)}{\partial \theta} = 0$$

Estimasi $\hat{\theta}$ dihitung dengan metode penskoran Fisher, yang biasa disebut dengan Informasi dari Fisher, misalnya pada model 2 parameter memenuhi rumus:

$$I(\theta) = \sum_{j=1}^n a_j^2 P_j(\theta) [1 - P_j(\theta)] \dots \dots \dots (2.20)$$

Iterasi dari penyelesaian penskoran Fisher yakni

$$\hat{\theta}_{i+1} = \hat{\theta}_i + I^{-1}(\hat{\theta}) \left(\frac{\partial \log L_i(\hat{\theta})}{\partial \theta} \right) \dots \dots \dots (2.21)$$

Kesalahan standar dari estimator ML merupakan kebalikan dari akar kuadrat nilai informasi pada $\hat{\theta}$

$$SE(\hat{\theta}) = \sqrt{\frac{1}{I(\hat{\theta})}} \dots \dots \dots (2.22)$$

Estimasi Bayes merupakan rerata dari distribusi posterior θ , setelah diberikan pola respons peserta hasil tes x_i . θ dapat didekati secara akurat dengan

$$\bar{\theta}_i = \frac{\sum_{k=1}^q X_k P(x_i | X_k) A(X_k)}{\sum_{k=1}^q P(x_i | X_k) A(X_k)} \dots \dots \dots (2.23)$$

Fungsi dari pola respons x_i sering disebut estimator dari Expected a Posteriori (EAP).

Ukuran ketepatan dari $\bar{\theta}_i$ merupakan standar deviasi posterior (Posterior standard deviation, PSD) yang didekati dengan

$$PSD(\bar{\theta}_i) = \frac{\sum_{k=1}^q (X_k - \bar{\theta}_i)^2 P(x_i | X_k) A(X_k)}{\sum_{k=1}^q P(x_i | X_k) A(X_k)} \dots \dots \dots (2.24)$$

Pembobotan ($A(X_k)$) pada formula tersebut didasarkan pada asumsi dari distribusi θ .

Estimasi dengan Bayes Modal mirip dengan Estimasi Bayes, namun dengan kesalahan rerata yang lebih besar. Estimasi ini sering disebut juga dengan Maximum a Posteriori (MAP). Estimator MAP merupakan nilai θ yang memaksimumkan

$$P(\theta|x_i) = \sum_{j=1}^n \{x_{ij} \log_e P_j(\theta) + (1 - x_{ij}) \log_e [1 - P_j(\theta)]\} + \log_e g(\theta) \dots\dots(2.25)$$

Dengan $g(\theta)$ fungsi kerapatan dari suatu distribusi populasi yang kontinu θ . Persamaan likelihoodnya

$$\sum_{j=1}^n \frac{x_{ij} - p_j(\theta)}{p_j(\theta)[1 - p_j(\theta)]} \cdot \frac{\partial p_j(\theta)}{\partial(\theta)} + \frac{\partial \log_e g(\theta)}{\partial \theta} = 0 \dots\dots\dots(2.26)$$

Analog dengan estimasi maksimum likelihood, MAP dihitung dengan penyekoran Fisher dengan menggunakan informasi porterior

$$J(\theta) = I(\theta) + \frac{\partial^2 \log_e g(\theta)}{\partial \theta^2}$$

Pada kasus 2PL, dengan distribusi normal dari kemampuan θ yang memiliki varians σ^2 , informasi posteriornya

$$I(\theta) = \sum_{j=1}^n a_j^2 p_j(\theta) [1 - P_j(\theta)] + \frac{1}{\sigma^2}$$

Dengan PSD dari estimasi MAP $\hat{\theta}$ didekati dengan

$$PSD(\hat{\theta}) = \sqrt{\frac{1}{I(\hat{\theta})}} \dots\dots\dots(2.27)$$

Menentukan Kecocokan Model

Ada 3 model yang dapat digunakan untuk melakukan analisis dengan menggunakan teori tes klasik, yakni model 1PL, 2PL, dan 3PL. Pengguna teori ini perlu memilih, data yang dianalisis apakah sesuai dengan salah satu dari ketiga model tersebut. Ada dua cara yang dapat digunakan untuk menentukan kecocokan model analisis yang akan digunakan, yakni dengan kecocokan model secara statistik dan dengan plot kurva karakteristik butir.

Pada pemilihan model secara statistik, dari ketiga model dibuat kecocokan butir berdasarkan nilai Khi-kuadrat. Kecocokan model ini dapat diketahui, dengan membandingkan khi-kuadrat hasil perhitungan dengan khi kuadrat tabel dengan derajat kebebasan tertentu. Butir dikatakan cocok dengan suatu model jika nilai khi-kuadrat

hitung tidak melebihi nilai khi-kuadrat tabel. Kecocokan dapat diketahui pula dari nilai probabilitas (signifikansi, sig). Jika nilai $\text{sig} < \alpha$, maka butir dikatakan tidak cocok dengan model.

Sebagai contoh, hasil analisis karakteristik perangkat tes seleksi masuk SMP Y Yogyakarta tahun 2002 (Nama dikodekan) (Heri Retnawati, 2003). Dari 15 butir perangkat tes yang dianalisis, ada 3 butir tidak cocok dengan model tiga parameter. Hasil selengkapnya disajikan pada Tabel 2.1.

Tabel 1
Sebagian Hasil Analisis Menggunakan Teori Respons Butir pada
Perangkat Tes Seleksi Masuk SLTPN Y Yogyakarta

Butir	a	b	c	χ^2	dk	χ^2 Kritis	Status
1.	0,466	-0,979	0,26	2,1	3	11,34	Cocok
3.	0,677	-0,427	0,209	14,1	3	11,34	Tidak cocok
4.	0,471	3,389	0,33	8,5	5	15,09	Cocok
5.	0,849	0,271	0,204	9,9	3	11,34	Cocok
6.	0,664	-0,24	0,259	2,7	3	11,34	Cocok
7.	0,749	2,734	0,383	1,5	5	15,09	Cocok
8.	0,406	0,799	0,284	2,2	4	13,28	Cocok
9.	0,497	0,349	0,31	1,3	3	11,34	Cocok
10.	0,658	1,254	0,214	3,1	4	13,28	Cocok
11.	0,846	-0,699	0,243	9,8	2	9,21	Tidak cocok
14.	0,747	3,529	0,292	7,5	5	15,09	Cocok
15.	0,66	3,021	0,303	2,4	5	15,09	Cocok
18.	0,845	1,163	0,17	7,3	4	13,28	Cocok
19.	1,111	1,126	0,157	4,7	4	13,28	Cocok
20.	0,637	1,154	0,31	0,3	4	13,28	Cocok

Pada suatu data, model logistik yang mempunyai butir yang cocok paling banyak dipilih sebagai model untuk analisis data. Misalnya pada kasus analisis ujian nasional SMP mata pelajaran matematika 2006. Data dianalisis dengan model 1PL, 2PL, dan 3PL. Kecocokan semua butir dengan model didaftar, hasil analisis tiap model disajikan dalam tabel. Perbandingan hasil analisis kecocokan butir dengan model 1PL, 2PL, dan 3PL pada data ujian nasional SMP mata pelajaran matematika 2006 disajikan pada Tabel 2.

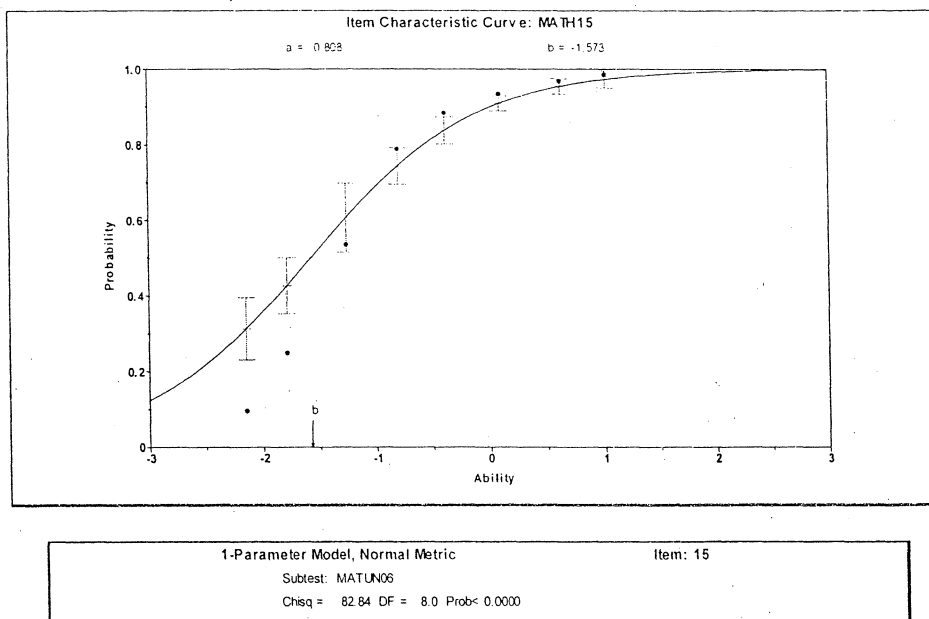
Tabel 2.2
Kecocokan Butir Perangkat UN Matematika SMP 2006
Berdasarkan Teori Respons Butir Model 1, 2, dan 3 Parameter

No. Butir	1P	2P	3P
1	Tidak Cocok	Tidak Cocok	Tidak Cocok
2	Tidak Cocok	Tidak Cocok	Tidak Cocok
3	Tidak Cocok	Tidak Cocok	Tidak Cocok
4	Tidak Cocok	Tidak Cocok	Tidak Cocok
5	Tidak Cocok	Cocok	Cocok
6	Tidak Cocok	Tidak Cocok	Tidak Cocok
7	Tidak Cocok	Tidak Cocok	Tidak Cocok
8	Tidak Cocok	Tidak Cocok	Tidak Cocok
9	Tidak Cocok	Tidak Cocok	Cocok
10	Cocok	Tidak Cocok	Tidak Cocok
11	Tidak Cocok	Cocok	Tidak Cocok
12	Tidak Cocok	Cocok	Cocok
13	Tidak Cocok	Cocok	Cocok
14	Tidak Cocok	Tidak Cocok	Tidak Cocok
15	Tidak Cocok	Cocok	Cocok
16	Tidak Cocok	Tidak Cocok	Tidak Cocok
17	Tidak Cocok	Tidak Cocok	Tidak Cocok
18	Tidak Cocok	Tidak Cocok	Cocok
19	Tidak Cocok	Tidak Cocok	Tidak Cocok
20	Tidak Cocok	Tidak Cocok	Cocok
21	Tidak Cocok	Tidak Cocok	Tidak Cocok
22	Tidak Cocok	Tidak Cocok	Tidak Cocok
23	Tidak Cocok	Tidak Cocok	Tidak Cocok
24	Tidak Cocok	Tidak Cocok	Cocok
25	Cocok	Cocok	Cocok
26	Tidak Cocok	Cocok	Cocok
27	Tidak Cocok	Tidak Cocok	Cocok
28	Tidak Cocok	Tidak Cocok	Cocok
29	Tidak Cocok	Tidak Cocok	Tidak Cocok
30	Tidak Cocok	Tidak Cocok	Tidak Cocok
Cocok Model	3	7	12
Tidak Cocok Model	27	23	18

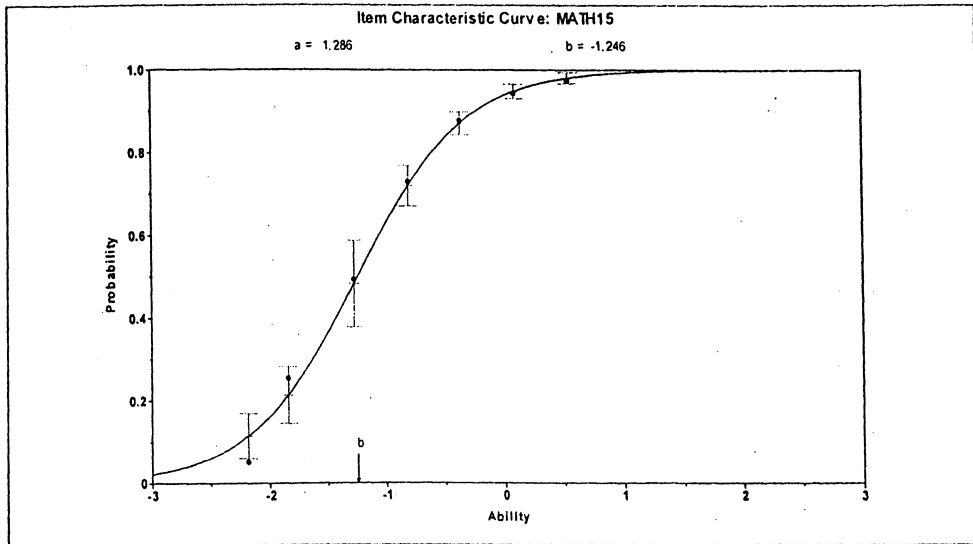
Berdasarkan hasil pada Tabel 3 tersebut, ternyata model yang menghasilkan butir yang cocok dengan model paling banyak adalah model 3 parameter. Ini berarti model 3 parameter merupakan model yang dapat dipilih untuk analisis butir.

Memperhatikan hasil analisis pada data ujian nasional SMP mata pelajaran matematika 2006, diperoleh bahwa hanya 12 dari 30 butir cocok dengan model 3P. Hal ini disebabkan respons data yang digunakan untuk analisis sangat banyak (lebih dari 3.000 peserta tes). Semakin banyak respons peserta tes yang digunakan, semakin besar perolehan nilai khi-kuadrat hitung. Semakin besar perolehan nilai khi-kuadrat hitung, semakin besar peluang menolak hipotesis butir cocok dianalisis dengan model logistik 3PL.

Cara kedua yang dapat dilakukan yakni dengan memuat plot kurva karakteristik. Plot ini dapat digambar dengan bantuan program BILOG Windows Version, yang disajikan pada Gambar 2.4, 2.5, dan 2.6 (Heri Retnawati, 2008). Dengan plot ini, dapat diketahui seberapa tepat distribusi data dibandingkan dengan modelnya. Sebagai contoh pada butir nomor 15. Mencermati gambar 8, kurva karakteristik butir 15 model 1P, distribusi data banyak yang letaknya jauh dibandingkan dengan model 2P (Gambar 2.5) dan 3P (gambar 2.6). Pada model 3P, distribusi data lebih mendekati 3PL dibandingkan 2PL. Hasil perbandingan butir pertama sampai butir ke-30 disajikan pada Tabel 2.3.

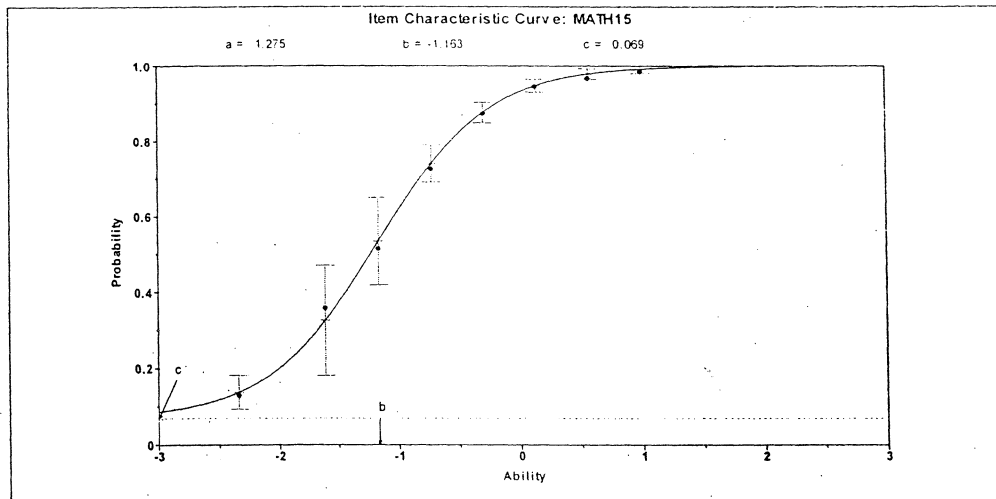


Gambar 2.4
 Plot Kurva Karakteristik Butir Nomor 15 Model 1 Parameter



2-Parameter Model, Normal Metric Item: 15
 Subtest: MATUN06
 Chisq = 10.01 DF = 8.0 Prob < 0.2640

Gambar 2.5
 Plot Kurva Karakteristik Butir Nomor 15 Model 2 Parameter



3-Parameter Model, Normal Metric Item: 15
 Subtest: MATUN06
 Chisq = 4.86 DF = 8.0 Prob < 0.7728

Gambar 2.6
 Plot Kurva Karakteristik Butir Nomor 15 Model 3 Parameter

Tabel 2.3
 Kesesuaian Butir Perangkat UN Matematika SMP 2006
 Berdasarkan Model 1, 2, dan 3 Parameter
 dengan Metode Plot Kurva Karakteristik Butir

No. Butir	Butir paling sesuai dengan Model		
	1P	2P	3P
1			√
2			√
3	√		
4		√	
5		√	
6			√
7	√		
8		√	
9	√		
10	√		
11		√	
12			√
13			√
14			√
15			√
16			√
17		√	
18			√
19			√
20			√
21			√
22			√
23			√
24			√
25		√	
26			√
27			√
28			√
29			√
30			√
Banyaknya butir yang cocok dengan model	4	6	20

Mencermati perbandingan pada ketiga plot kurva karakteristik pada Tabel 2.3 tersebut ini, dapat diperoleh bahwa model 3P merupakan model yang paling baik dibandingkan dengan kedua model lainnya, yakni model 2P dan model 1P. Berdasarkan pertimbangan dengan kedua cara ini, baik dengan statistik maupun dengan menggunakan plot, ditetapkan pada analisis penelitian ini dilakukan dengan model 3P. Selanjutnya estimasi parameter dan kemampuan dilakukan dengan model 3PL yang hasilnya disajikan pada Tabel 2.4. Contoh hasil analisis selanjutnya disajikan pada tabel berikut, dengan diberi interpretasi butir baik atau tidak baik sesuai klasifikasi parameter. Butir baik selanjutnya dapat disimpan dalam bank butir sesuai tujuan instrument yang memuat tersebut dikembangkan.

Tabel 2.4
Karakteristik Perangkat Tes UN Matematika SMP 2006
Berdasarkan Teori Respons Butir Unidimensi Model 3 Parameter

Butir	Materi	a	b	c	Keterangan
1	Persentase (soal cerita)	1,229	-1,242	0,017	Baik
2	Diagram Venn	0,996	-1,089	0,018	Baik
3	Persentase	0,667	-0,023	0,045	Baik
4	HP bil bulat	0,526	0,400	0,035	Baik
5	Jaring-jaring kubus	0,930	-2,552	0,144	Kurang baik ($b < -2.0$)
6	Simetri lipat	0,520	-1,927	0,070	Baik
7	Sudut segitiga	0,887	-1,034	0,500	Kurang baik ($c > 0,25$)
8	Pemetaan	1,030	-1,048	0,026	Baik
9	Akar dan pangkat	1,083	-1,311	0,033	Baik
10	Sifat garis sejajar	0,677	-0,181	0,036	Baik
11	Keliling belah ketupat	1,321	-0,709	0,022	Baik
12	Luas Jajar genjang	1,104	-0,517	0,026	Baik
13	Perbandingan (soal cerita)	1,175	-1,999	0,159	Baik
14	Persamaan garis lurus	0,762	0,018	0,046	Baik
15	SPL (soal cerita)	1,150	-1,296	0,041	Baik
16	Median data	0,824	-0,397	0,019	Baik
17	Volume limas	0,868	-0,182	0,014	Baik
18	Luas permukaan prisma	1,135	-0,656	0,105	Baik
19	Refleksi	1,231	-0,640	0,255	Kurang baik ($c > 0,25$)
20	Dilatasi	0,803	-0,741	0,094	Baik
21	Perbandingan segitiga	2,847	0,659	0,352	Kurang baik ($c > 0,25$)
22	Segitiga kongruen	0,658	-1,250	0,038	Baik
23	Juring lingkaran	1,076	0,051	0,359	Kurang baik ($c > 0,25$)
24	Persekutuan lingkaran	1,114	-0,667	0,136	Baik
25	Suku dan faktor	0,880	-0,792	0,049	Baik
26	Fungsi kuadrat	1,064	-0,502	0,116	Baik

27	Phytagoras dan luas segitiga	1,247	-0,690	0,292	Kurang baik ($c > 0,25$)
28	Barisan dan deret	1,328	-1,522	0,035	Baik
29	Trigonometri	0,717	-0,265	0,500	Kurang baik ($c > 0,25$)
30	Logaritma	1,170	0,045	0,500	Kurang baik ($c > 0,25$)

Untuk melakukan analisis, ada beberapa software yang dapat digunakan. Untuk model Rasch (model 1PL), diantaranya dapat digunakan software BIGSTEPS, WINSTEPS, QUEST, CONQUEST. Untuk model 2PL dan 3PL dapat digunakan software BILOGMG, MULTILOG, PARSCALE, dan MPLUS.

* * *

Daftar Pustaka

- Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.
- Heri Retnawati. (2003). Keberfungsian butir diferensial pada perangkat tes seleksi masuk SMP. Tesis. Universitas Negeri Yogyakarta, tidak dipublikasikan.
- Heri Retnawati (2008). Estimasi efisiensi relative tes berdasarkan teori tes klasik dan teori respons butir. *Disertasi*. Universitas Negeri Yogyakarta, tidak dipublikasikan.
- Hosmer, D.W. dan Lemeshow, S. (1989). *Applied Logistic Regressions*. New York : John Willwy and Sons.
- Hullin, C. L., et al. (1983). *Item response theory : Application to psychological measurement*. Homewood, IL : Dow Jones-Irwin.
- Keeves, J.P. dan Alagumalai, S. (1999). New approaches to measurement. Dalam Masters, G.N. dan Keeves, J.P. (Eds). *Advances in measurement in educational research and assesment*. Amsterdam : Pergamon.
- Van der Linden, W.J. dan Hambleton, R.K. (1997). Item response theory: brief history, common models and extentions. Dalam Van der Linden, W.J. dan Hambleton, R.K. (Eds). *Handbook of item response theory*. New York : Springer.

BAB 3

TEORI RESPONS BUTIR

POLITOMUS



Selain model respons butir dikotomi, ada model lain yang dapat digunakan untuk menskor respons peserta terhadap suatu butir tes, yakni model politomi. Model-model politomi pada teori respons butir antara lain *nominal resons model (NRM)*, *rating scale model (RSM)*, *partial credit model (PCM)*, *graded respons model (GRM)* dan *generalized partial credit model (GPCM)* (Van der Linden & Hambleton, 1997).

Model respons butir politomous dapat dikategorikan menjadi model respons butir nominal dan ordinal, tergantung pada asumsi karakteristik tentang data. Model respons butir nominal dapat diterapkan pada butir yang mempunyai alternatif jawaban yang tidak terurut (*ordered*) dan adanya berbagai tingkat kemampuan yang diukur. Pada model respons ordinal terjadi pada butir yang dapat diskor ke dalam banyaknya kategori tertentu yang tersusun dalam jawaban. Skala Likert diskor berdasarkan pedoman penskoran kategori respons terurut, yang merupakan penskoran ordinal. Butir-butir tes matematika dapat diskor menggunakan sistem parsial kredit, langkah-langkah menuju jawaban benar dihargai sebagai penskoran ordinal. Model penskoran yang pang sering dipakai ahli yakni GRM, PCM, dan GPCM.

Contoh model penskoran untuk GRM misalnya pada angket menggunakan skala Likert. Pada skala Likert, peserta dapat menjawab Sangat Setuju (SS), Setuju (S), Netral

(N), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Penskoran dibedakan untuk pernyataan positif dan pernyataan negatif, seperti disajikan pada Tabel 3.1.

Tabel 3.1. Contoh penskoran pada skala Likert

Pernyataan	Jenis	Skoring				
		STS	TS	N	S	SS
Mencari berbagai sumber informasi baik buku, majalah, dan internet jika ada hal yang ingin saya ketahui	Positif	1	2	3	4	5
Ketika guru menjelaskan dan ada yang hal yang belum saya pahami, saya diam saja dan menunggu teman biar menanyakannya.	Negatif	5	4	3	2	1

Pada kasus tersebut, pendapat responden diberi skor berjenjang yang menunjukkan tingkatan, mulai dari yang terendah ke yang tertinggi.

Penskoran parsial biasanya dilakukan pada instrumen yang ada bagian-bagiannya. Misalnya pada instrumen untuk mengobservasi kemandirian anak menggosok gigi. Untuk menggosok gigi, diperlukan paling tidak 7 tahap atau 7 bagian sebagai berikut.

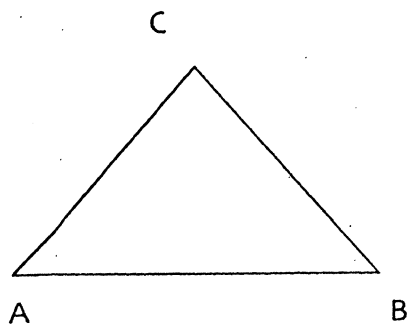
Menggosok gigi

- Mengambil sikat
- Mengoleskan pasta gigi ke bulu sikat gigi
- Berkumur-kumur
- Menggosok gigi dengan sikat
- Berkumur-kumur
- Mengembalikan sikat ke tempatnya
- Mengembalikan pasta gigi ke tempatnya

Pada instrumen tersebut, responden yang diamati diberikan skor untuk tiap langkah yang dilakukannya. Responden kadang tidak melakukan semua tahap, dan bisa jadi tidak berurutan namun dilakukan. Langkah yang dilakukan diskor 1, yang tidak diskor nol. Total skor yang diperoleh merupakan penskoran dengan model parsial.

Contoh lain dari bentuk penskoran politomi jenis parsial adalah penskoran pada tes jenis uraian. Untuk uraian, penskoran dilakukan dengan melihat tahap-tahap peserta tes dalam menyelesaikan soal.

Sebagai contoh butir soal berikut.



Sebuah kolam berbentuk segitiga samakaki seperti yang digambarkan pada gambar di samping. Jika panjang AB 12 m, dan panjang AC 10 m. Jika biaya untuk membuat kolam per meter Rp. 150.000,-, berapakah biaya total untuk membuat kolam tersebut?

Agar penilaian menjadi lebih objektif, penyusun instrumen perlu membuat suatu rubric pedoman penskoran. Sebagai contoh rubriknya disajikan pada Tabel 3.2 sebagai berikut.

Tabel 3.2 Contoh penskoran pada soal pilihan ganda

Langkah	Skor
Menghitung tinggi segitiga	1
Menghitung luas segitiga	1
Menghitung biaya pembuatan kolam	1
Total	3

Model penskoran lainnya yakni model penskoran nominal (nominal response model) dan penskoran pilihan ganda. Contoh penskoran dengan model nominal misalnya pada kasus pilihan presiden. Pada buku ini hanya dibahas GRM dan PCM dan perluasannya.

1. Graded Respons Model (GRM)

Respons peserta terhadap butir j dengan model GRM dikategorikan menjadi $m+1$ skor kategori terurut, $k=0,1,2,\dots,m$ dengan m merupakan banyaknya langkah dalam menyelesaikan dengan benar butir j , dan indeks kesukaran dalam setiap langkah juga terurut. Hubungan parameter butir dan kemampuan peserta dalam GRM untuk kasus

homogen (aj sama dalam setiap langkah) dapat dinyatakan oleh Muraki & Bock (1997:7) sebagai berikut.

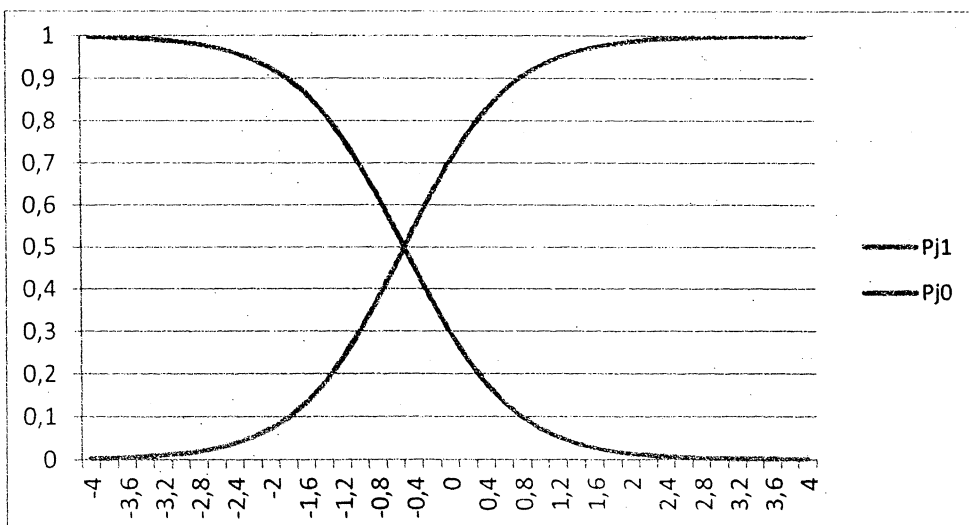
$$P_{jk}(\theta) = P_{jk}^*(\theta) - P_{j,k+1}^*(\theta) \dots\dots\dots(3.1)$$

$$P_{jk}^*(\theta) = \frac{\exp[Da_j(\theta - b_{jk})]}{1 + \exp[Da_j(\theta - b_{jk})]} \dots\dots\dots(3.2)$$

Dengan $P_{j0}^*(\theta) = 1$ dan $P_{j,m+1}^*(\theta) = 0$

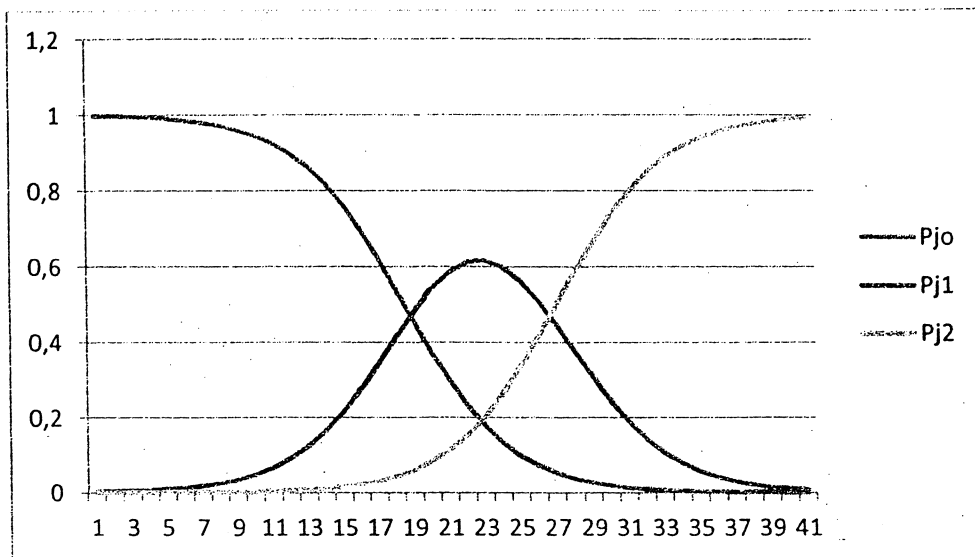
- a_j : indeks daya beda butir j
- θ : kemampuan peserta,
- b_{jk} : indeks kesukaran kategori k butir j
- $P_{jk}(\theta)$: probabilitas peserta berkemampuan θ yang memperoleh skor kategori k pada butir j
- $P_{jk}^*(\theta)$: probabilitas peserta berkemampuan θ yang memperoleh skor kategori k atau lebih pada butir j
- D : faktor skala

Hubungan antara kemampuan dengan peluang menjawab benar digambarkan dengan fungsi respons kategori (*Categorical Response Function, CRF*) (du Toit, 2003). Pada gambar 3.1 disajikan CRF untuk GRM.

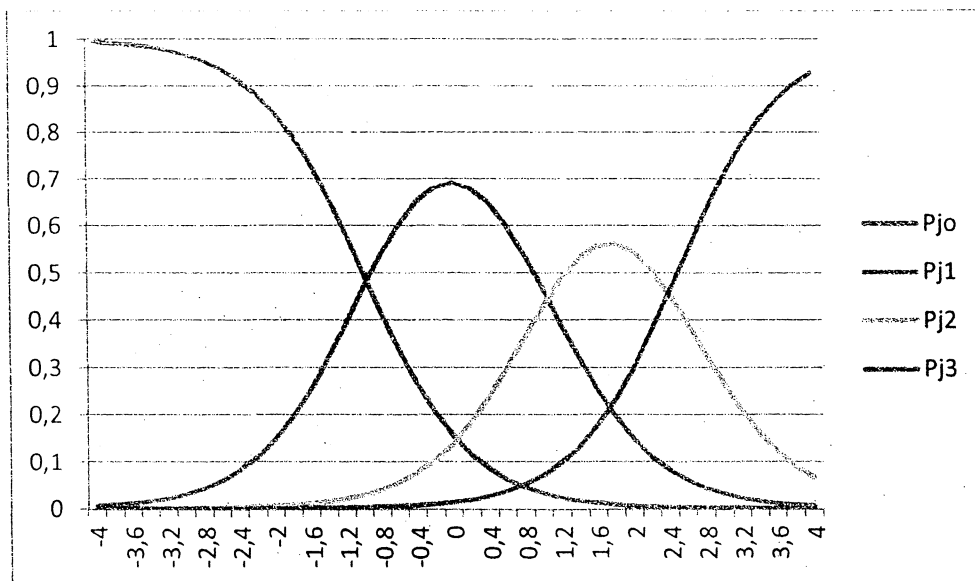


Gambar 3.1. CRF 2 kategori (a=1, b=-0,5) dengan GRM

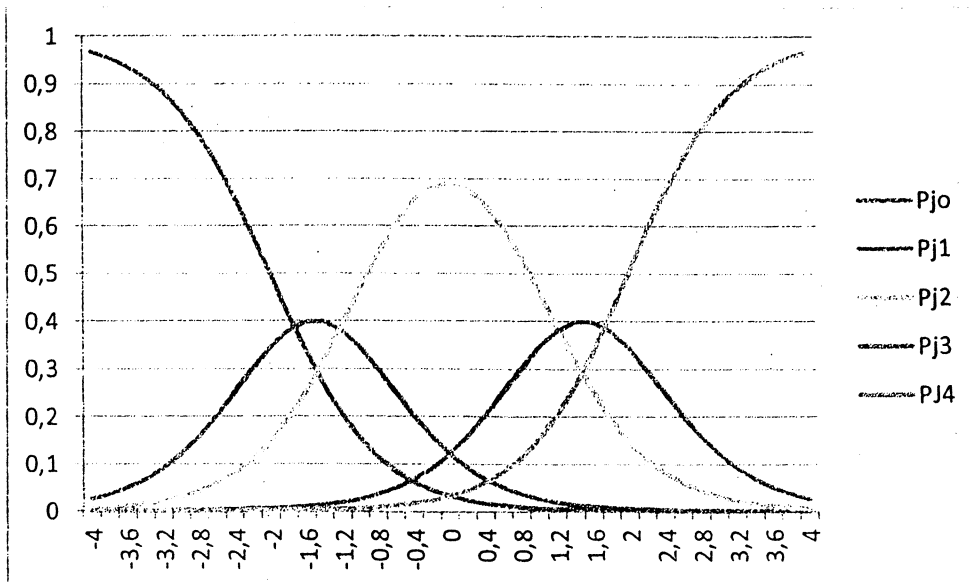
Mencermati gambar 3.1 tersebut, dapat diperoleh bahwa P_{ji} sama dengan model logistik dikotomi dengan $a=1$, $b=-0,5$, dan $P_{j1} = 1 - P_{j0}$. Untuk 3, 4, 5 kategori, CRF disajikan pada Gambar 3.2, 3.3, 3.4, dan 3.5.



Gambar 3.2. CRF 3 kategori ($a=1$, $b_1=-0,5$, $b_2=1,2$) dengan GRM



Gambar 3.3. CRF 4 kategori ($a=1$, $b_1=-1$, $b_2=1$, $b_3=2,5$) dengan GRM



Gambar 3.4. CRF 5 kategori ($a=1, b_1=-2, b_2=-1, b_3=1, b_4=2$) dengan GRM

2. Partial Credit Model (PCM)

Pada awal perkembangan teori respons butir politomus, model yang lebih dikenal yakni perluasan dari model Rasch yang disebut dengan *Partial Credit Model* (PCM). PCM merupakan model penskoran politomus yang merupakan perluasan dari model Rasch pada data dikotomi. Asumsi pada PCM yakni setiap butir mempunyai daya beda yang sama. PCM mempunyai kemiripan dengan *Graded Response Model* (GRM) pada butir yang diskor dalam kategori berjenjang, namun indeks kesukaran dalam setiap langkah tidak perlu terurut, suatu langkah dapat lebih sukar dibandingkan langkah berikutnya.

Bentuk umum PCM menurut Muraki & Bock (1997:16) sebagai berikut.

$$P_{jk}(\theta) = \frac{\exp \sum_{v=0}^k (\theta - b_{jv})}{\sum_{h=0}^m \exp \sum_{v=0}^h (\theta - b_{jv})}, k=0,1,2,\dots,m \dots\dots\dots(3.3)$$

Dengan

$P_{jk}(\theta)$ = probabilitas peserta berkemampuan θ memperoleh skor kategori k pada butir j,

θ : kemampuan peserta,

$m+1$: banyaknya kategori butir j ,

b_{jk} : indeks kesukaran kategori k butir j

$$\sum_{h=0}^k (\theta - b_{jh}) \equiv 0 \quad \text{dan} \quad \sum_{h=0}^h (\theta - b_{jh}) \equiv \sum_{h=1}^h (\theta - b_{jh}) \dots\dots\dots(3.4)$$

Skor kategori pada PCM menunjukkan banyaknya langkah untuk menyelesaikan dengan benar butir tersebut. Skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih besar daripada skor kategori yang lebih rendah. Pada PCM, jika suatu butir memiliki dua kategori, maka persamaan 2 menjadi persamaan model Rasch, seperti persamaan yang dinyatakan oleh Hambleton, Swaminathan (1985), dan juga diperkuat oleh Hambleton, Swaminathan, dan Roger (1991). Sebagai akibat dari hal ini, PCM dapat diterapkan pada butir politomus dan dikotomus.

Pengembangan lebih lanjut penskoran politomus adalah *Generalized Partial Credit Model* (GPCM). GPCM menurut Muraki (1999) merupakan bentuk umum dari PCM, yang dinyatakan dalam bentuk matematis, yang disebut sebagai fungsi respons kategori butir sebagai berikut.

$$P_{jh}(\theta) = \frac{\exp \sum_{v=0}^h Z_{jv}(\theta)}{\sum_{e=0}^{m_j} \exp \left[\sum_{v=0}^e Z_{jv}(\theta) \right]}, \quad k=0,1,2,\dots,m_j \dots\dots\dots(3.5)$$

dan

$$Z_{jh}(\theta) = Da_j(\theta - b_{jh}) = Da_j(\theta - b_j + d_h), \quad b_{j0} = 0 \dots\dots\dots(3.6)$$

Dengan

$P_{jk}(\theta)$: probabilitas peserta berkemampuan θ memperoleh skor kategori k pada butir j ,

θ : kemampuan peserta,

a_j : indeks daya beda butir j ,

b_{jh} : indeks kesukaran kategori k butir j ,

b_j : indeks kesukaran lokasi butir j (parameter butir lokasi)

d_k : parameter kategori k ,

m_j+1 : banyaknya kategori butir j , dan

D : faktor skala ($D=1.7$)

Parameter b_{jk} oleh Master dinamai dengan parameter tahap butir. Parameter ini merupakan titik potong antara kurva $P_{jk}(\theta)$ dengan $P_{jk-1}(\theta)$. Kedua kurva hanya berpotongan di satu titik pada skala θ (van der Linden & Hambleton, 1997).

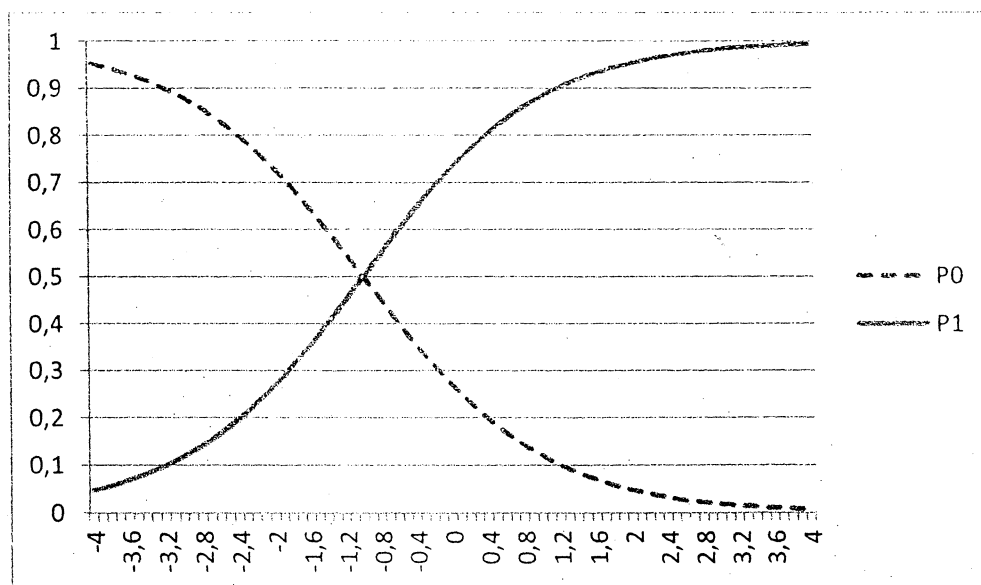
Jika $\theta = b_{jk}$, maka $P_{jk}(\theta) = P_{jk-1}(\theta)$

Jika $\theta > b_{jk}$, maka $P_{jk}(\theta) > P_{jk-1}(\theta)$

Jika $\theta < b_{jk}$, maka $P_{jk}(\theta) < P_{jk-1}(\theta)$, $K=1,2,3,\dots,m_j$ (3.7)

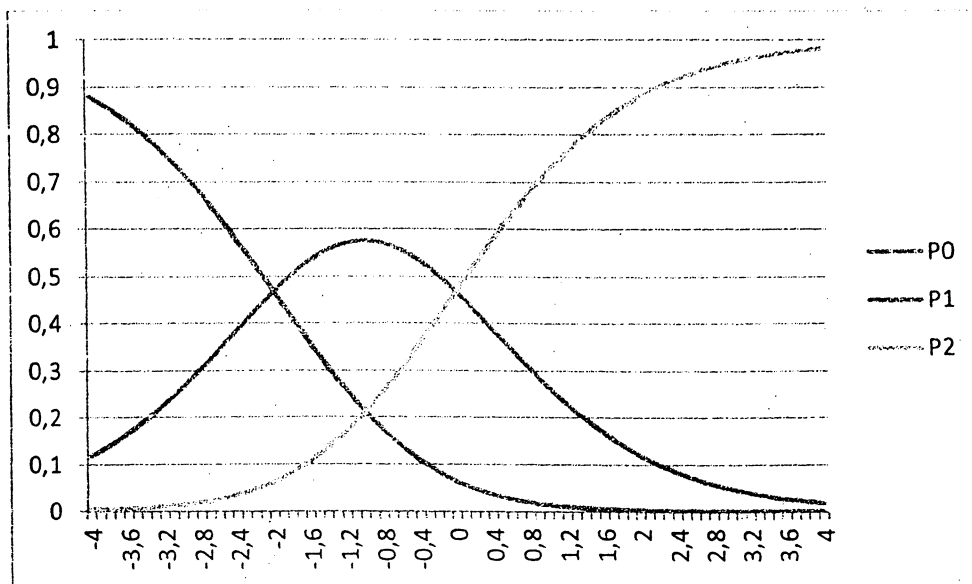
GPCM diformulasikan berdasarkan asumsi bahwa setiap probabilitas memilih kategori ke-k melampaui kategori ke-(k-1) dibangun oleh model dikotomi. P_{jk} merupakan probabilitas khusus memilih kategori ke-k dari $m_j + 1$ kategori. Hubungan probabilitas menjawab benar untuk tiap kemampuan θ disajikan dalam grafik *Categorical Response Function* (CRF) (du Toit, 2003). Grafik CRF pada 2, 3, 4, dan 5 kategori disajikan pada Gambar 4.5, 4.6, 4.7, dan 4.8.

Pada dikotomis model, ada 2 kategori yaitu 1 dan 0. Untuk daya pembeda (a) sebesar 1,0 dan tingkat kesulitan pada kategori menjawab 1 sebesar -1,0, disajikan pada Gambar 4.5.



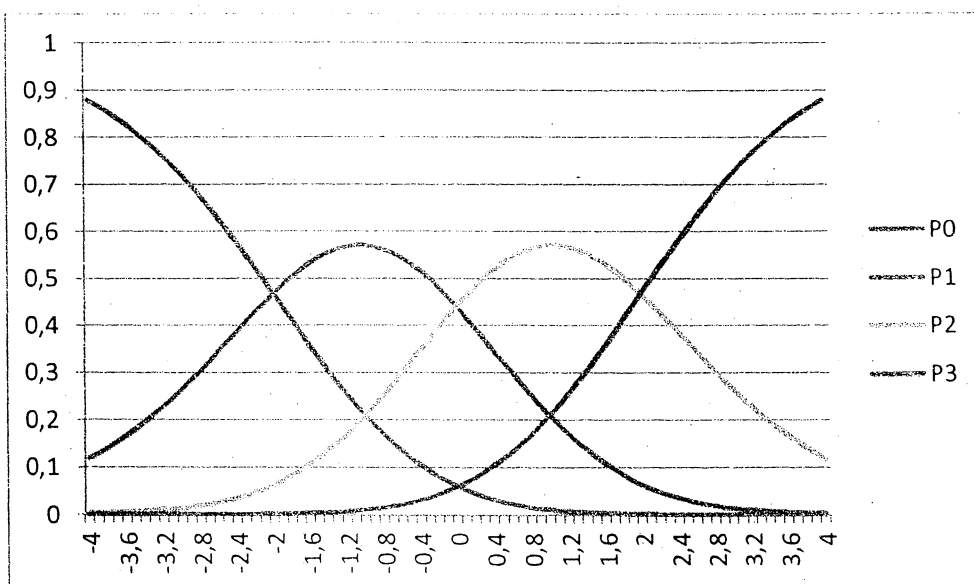
Gambar 4.5. Grafik CRF pada 2 kategori dengan GPCM

Pada dikotomis model, ada 3 kategori yaitu 0, 1 dan 2. Untuk daya pembeda (a) sebesar 1,0 dan tingkat kesulitan pada kategori menjawab 1 sebesar -2,0 dan 0,0, disajikan pada Gambar 2.



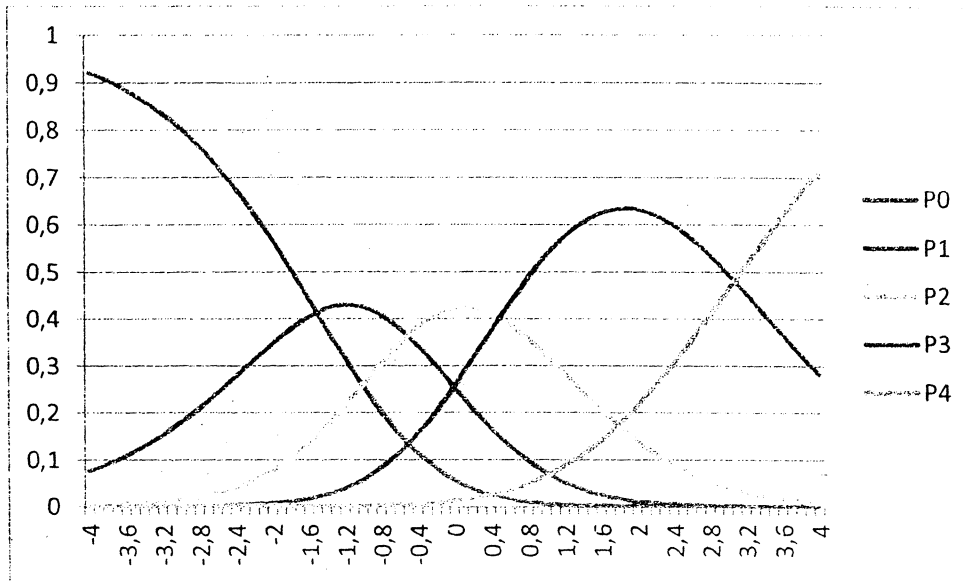
Gambar 4.6. Grafik CRF pada 3 kategori dengan GPCM

Pada Gambar 4.7 disajikan model politomis dengan 4 kategori yaitu 0,1,2 dan 3. Untuk daya pembeda (a) sebesar 1,0 dan tingkat kesulitan pada kategori menjawab -2,0, 0,0 dan 2,0.



Gambar 4.7. Grafik CRF pada 4 kategori dengan GPCM

Pada Gambar 4.8 disajikan model politomus dengan 4 kategori yaitu 0,1,2 dan 3. Untuk daya pembeda (a) sebesar 1,0 dan tingkat kesulitan pada kategori menjawab -1,5, -0,5, 0,5 dan 1,5.



Gambar 4.8 Grafik CRF pada 5 Kategori dengan GPCM

Seperti halnya pada teori respons butir, pada model politomi dikenal dengan nilai fungsi karakteristik butir yang kemudian biasanya digunakan untuk menggambarkan kurva karakteristik butir (Item Characteristic Curve, ICC).

$$E(X|\theta) = \sum_{i=1}^n \sum_{x=0}^{m_i} x P_{ix}(\theta) \dots\dots\dots(3.8)$$

Agar informasi yang diperoleh berguna dalam penskoran tes, parameter butir perlu diestimasi. Estimasi parameter butir dan mengecek kecocokan model sering disebut sebagai kalibrasi butir. Kalibrasi ini dapat dilakukan jika data respons peserta terhadap tes telah diperoleh (du Toit, 2003). Paling tidak ada 2 pendekatan yang dapat digunakan untuk estimasi parameter butir atau melakukan kalibrasi butir, yakni estimasi *Marginal Maximum Likelihood* (MML) dan estimasi *Marginal Maximum A Posteriori* (MMAP).

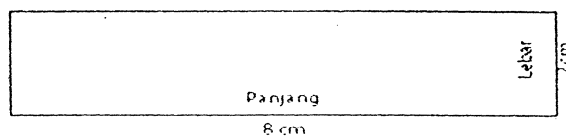
MML merupakan metode yang diyakini efisien untuk semua model respons butir dan untuk tes yang panjang maupun yang pendek. MML mengasumsikan adanya respons yang berbeda dari kemampuan θ yang sama. Untuk mengetahui parameter butir, metode yang terkenal yakni metode Bock & Lieberman, yang kemudian dirumuskan kembali oleh Bock & Aitkin tahun 1981 untuk sampel besar (Muraki, 1997). Metode ini terdiri dari 2 langkah, yakni langkah estimasi dan langkah maksimasi. Pada langkah estimasi, frekuensi

harapan provisional $\overline{r_{ihf}}$ dan ukuran sampel harapan provisional $\overline{N_f}$ dihitung. Kemudian pada langkah maksimasi, diestimasi Marginal Maximum Likelihood (MML) dengan metode penskoran Fisher.

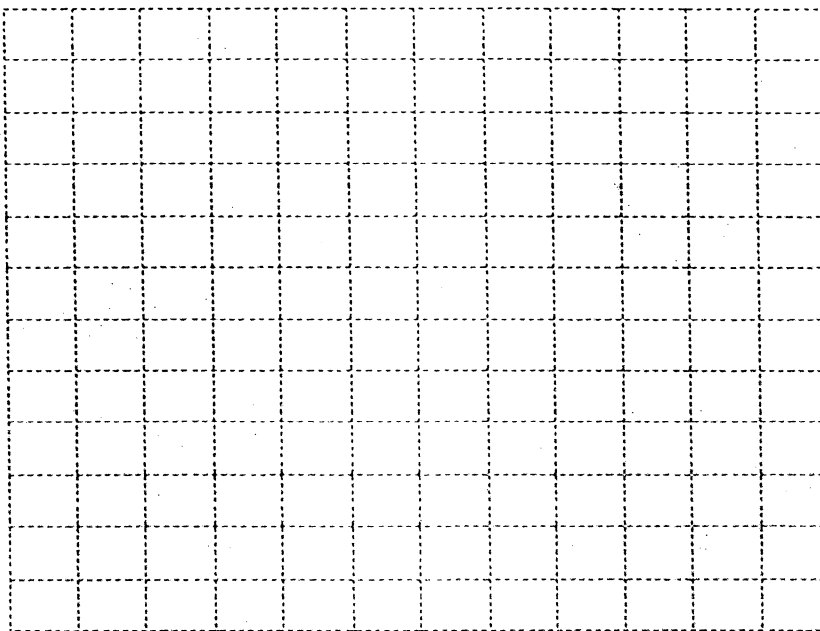
Program yang digunakan untuk mengestimasi parameter butir dan kemampuan dengan penskoran politomi diantaranya Parscale dari SSI (Muraki & Bock, 1997), Multilog, Winsteps, Quest, Conquest, dan lain-lain. Untuk dapat menggunakan program ini, ada 2 hal yang perlu menjadi perhatian yakni input data dan sintaks analisis, yang masing-masing program memiliki bahasa yang unik.

Contoh hasil analisis pada 5 butir tes yang diskor secara politomi misalnya pada data TIMSS, untuk soal jenis *constructed response* (CR). Rubrik untuk tiap butir yakni salah (skor 0), betul sebagian (skor 1), dan betul keseluruhan (skor 2). Misalnya untuk butir matematika M022234A.

Naskah butir soal M022234A (sudah direlease) sebagai berikut.



- A. Pada kertas berpetak di bawah ini, gambarlah sebuah persegi panjang yang panjangnya sama dengan tiga perempat kali panjang persegi panjang di atas, dan lebarnya sama dengan dua setengah kali lebar persegi panjang di atas. Tuliskanlah panjang dan lebar persegi panjang yang kamu gambar dalam cm. Setiap kotak berukuran 1 cm x 1 cm.



- B. Berapa perbandingan luas persegi panjang di atas dengan persegi panjang yang kamu gambar?

Setelah dilakukan analisis dengan Parscale (Muraki & Bock, 1997) menggunakan model GPCM, diperoleh parameter butir yang disajikan pada Tabel 3.4 sebagai berikut.

Tabel 3.4. Hasil analisis parameter butir pada data TIMSS 2007 mapel matematika

No.	Item	Content	Topic	Item		Slope (a _i)		Location (b _i)		Step 1 (d _{1i})		Step 2 (d _{2i})	
				Applying	CR	0,53	0,01	1,59	0,03	-2,18	0,07	2,18	0,08
77	M022232	Number	Fractions and Decimals	Applying	CR	0,53	0,01	1,59	0,03	-2,18	0,07	2,18	0,08
78	M022234A	Geometry	Geometric Shapes	Applying	CR	0,80	0,01	0,77	0,01	-0,63	0,03	0,63	0,03
79	M022234B	Number	Ratio, Proportion and Percent	Applying	CR	0,90	0,02	1,08	0,01	-1,48	0,05	1,48	0,05
80	M042220	Data and Chance	Data Organization and Representation	Applying	CR	0,69	0,02	0,44	0,03	-1,52	0,08	1,52	0,09
81	M042304B	Number	Ratio, Proportion and Percent	Applying	CR	0,97	0,04	0,67	0,03	0,41	0,03	-0,41	0,04
82	M042304D	Number	Whole Numbers	Reasoning	CR	0,58	0,02	0,23	0,03	-1,41	0,08	1,41	0,08
83	M042303B	Data and Chance	Data Interpretation	Reasoning	CR	0,37	0,02	0,79	0,06	-0,05	0,08	0,05	0,10
84	M032640	Algebra	Patterns	Reasoning	CR	0,61	0,03	1,56	0,07	-0,80	0,10	0,80	0,12
85	M032755	Number	Ratio, Proportion and Percent	Reasoning	CR	1,10	0,05	1,22	0,04	-0,29	0,06	0,29	0,07
86	M032753A	Data and Chance	Data Interpretation	Reasoning	CR	1,12	0,05	0,74	0,03	-0,29	0,05	0,29	0,06
87	M032753B	Data and Chance	Data Interpretation	Reasoning	CR	1,21	0,06	0,91	0,03	-0,03	0,04	0,03	0,05
88	M042059	Number	Ratio, Proportion and Percent	Knowing	CR	0,77	0,03	0,08	0,02	-0,18	0,05	0,18	0,05
89	M042207	Data and Chance	Data Organization and Representation	Applying	CR	0,44	0,01	0,02	0,03	-2,99	0,15	2,99	0,14
90	M032695	Data and Chance	Data Organization and Representation	Applying	CR	0,55	0,01	0,20	0,02	-1,06	0,05	1,06	0,05
91	M032683	Algebra	Algebraic Expression	Knowing	CR	0,49	0,01	0,87	0,03	-1,60	0,06	1,60	0,07
92	M032757	Algebra	Patterns	Reasoning	CR	0,48	0,02	0,19	0,04	-2,27	0,15	2,27	0,15
93	M032760A	Algebra	Patterns	Reasoning	CR	0,81	0,03	0,67	0,03	-1,39	0,10	1,39	0,11
94	M032761	Algebra	Algebraic Expression	Reasoning	CR	1,05	0,05	1,25	0,04	-0,41	0,06	0,41	0,08
95	M032692	Geometry	Geometric Shapes	Reasoning	CR	0,69	0,03	0,98	0,04	-0,99	0,09	0,99	0,10

Selain parameter butir, diperoleh pula parameter kemampuan peserta, pada skala (-4,4) dan dapat pula diperoleh plot kurva karakteristik butir, kurva fungsi kategori respons, nilai fungsi informasi, dan kesalahan pengukuran.



Daftar Pustaka

- Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: SSI.
- Muraki, E. (1999). New approaches to measurement. Dalam Masters, G.N. dan Keeves, J.P.(Eds). *Advances in measurement in educational research and assesment*. Amsterdam : Pergamon.
- Muraki,E., & Bock, R.D. (1997). *Parscale 3: IRT based test scoring and item analysis for graded items and rating scales*. Chicago: Scintific Software Inc.
- Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.

BAB 4

TEORI RESPONS BUTIR MULTIDIMENSI



Seperti halnya teori respons butir unidimensi, pada model teori respons butir multidimensi data dapat berupa butir skor dikotomi atau politomi. Matriks data disusun sedemikian rupa, dengan x_{ij} menyatakan elemennya pada baris ke- i dan kolom ke- j . Butir dinyatakan dalam i ($i=1, \dots, n$) dan peserta dinyatakan dalam j ($j=1, \dots, N$).

Dalam menyusun matriks data, ada asumsi yang harus diperhatikan (Reckase, 1997), yakni :

- a. Semakin tinggi kemampuan peserta tes, semakin besar probabilitas menjawab benar peserta tes terhadap butir soal (asumsi kemonotonan).
- b. Fungsi probabilitas menjawab benar bersifat smooth (turunan fungsinya terdefiniskan).
- c. Probabilitas kombinasi respons dapat ditentukan dengan hasil probabilitas respons individual ketika probabilitas dihitung kondisional pada titik dalam ruang yang didefinisikan oleh konstruk hipotetik (asumsi independensi lokal). Berdasarkan hal ini, asumsi yang digunakan untuk menyusun matriks data yakni asumsi kemonotonan, memiliki turunan fungsi, dan independensi lokal.

Pada teori respons butir multidimensi (*multidimensional item response theory*, MIRT) dikenal dua model, yakni *compensatory* dan *noncompensatory*. Menurut Ansley dan Forsyth (Spray, Davey, Reckase, et al., 1990), model *compensatory* membolehkan

kemampuan tinggi pada salah satu dimensi memperoleh kompensasi pada kemampuan rendah pada dimensi lain dalam kaitannya dengan probabilitas menjawab benar. Sebaliknya, pada model *noncompensatory* tidak membolehkan kemampuan tinggi pada salah satu dimensi memperoleh kompensasi pada kemampuan rendah pada dimensi lainnya. Untuk model *compensatory* pada kasus butir dua dimensi, seorang peserta tes dengan kemampuan sangat rendah pada satu dimensi dan kemampuannya sangat tinggi pada dimensi lain dapat menjawab butir tes dengan benar.

Ada dua tipe model *compensatory*, yakni model MIRT logistik (Reckase, 1997) dan model *ogive normal* dari Samejima dengan menyatakan kombinasi linear dari kemampuan multidimensi dalam pangkat pada rumus probabilitas menjawab benar. Dalam model linear ini, rendahnya satu atau lebih kemampuan dapat dikompensasikan pada dimensi lainnya. Karena kompensasi merupakan karakteristik kombinasi linear, maka model ini diberi nama dengan model MIRT linear (Spray, Davey, Reckase, et al., 1990; Bolt & Lall, 2003). Model MIRT logistik linear dapat ditulis sebagai:

$$P_i(\theta_j) = c_i + (1-c_i) \frac{e^{[\sum_{m=1}^k f_{ijm}] + d_i}}{(1 + e^{[\sum_{m=1}^k f_{ijm}] + d_i})} \dots\dots\dots(4.1)$$

Keterangan :

- $P_i(\theta_j)$: peluang peserta ke-j dengan kemampuan θ_j menjawab benar butir i
- θ_j : vektor kemampuan orang ke j
- f_{ijm} : nilai yang besarnya samadengan $a_{im}' \theta_{jm}$
- a_{im} : diskriminasi untuk butir ke-i pada dimensi ke-m
- θ_{jm} : elemen ke-m dari vektor kemampuan orang ke j (θ_j)
- c_i : parameter *pseudo-guessing* butir ke-i
- d_i : tingkat kesulitan butir ke-i

Senada dengan itu, Kirisci, Hsi, & Yu (2001) menuliskan persamaan (19) sebagai berikut.

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{(1 + e^{-1.7(a_i'\theta_j - d_i)})} \dots\dots\dots(4.2)$$

Keterangan :

- $P_i(\theta_j)$: peluang peserta ke-j dengan kemampuan θ_j menjawab benar butir i
- θ_j : vektor kemampuan orang ke j
- a_i : diskriminasi untuk butir ke-i pada dimensi ke-m
- θ_j : vektor kemampuan orang ke j (θ_j)
- c_i : parameter *pseudo-guessing* butir ke-i
- d_i : tingkat kesulitan butir ke-i

Di lain pihak, model MIRT *noncompensatory* dideskripsikan sebagai probabilitas dari respons yang menguntungkan pada hasil kali dari fungsi kemampuan sebanyak k dimensi dan karakteristik butir. Model MIRT logistik tipe *noncompensatory* dapat ditulis sebagai :

$$P_i(\theta_j) = c_i + (1-c_i) \prod_{m=1}^k \frac{e^{f_{ijm}}}{(1 + e^{f_{ijm}})} \dots\dots\dots(4.3)$$

Dengan $e^{f_{ijm}} = [a_{im}(\theta_{jm} - b_{im})]$ dengan b_{im} merupakan parameter butir ke-i pada dimensi ke-m. Terkait dengan bentuknya yang merupakan hasil perkalian, model ini sering pula dinamai dengan model multiplikatif.

Mengingat pada penelitian ini lebih difokuskan pada MIRT model *compensatory*, maka hanya model linear ini saja yang akan dibahas. Seperti halnya pada teori respons butir model 3 parameter, parameter-parameter model ini meliputi parameter peserta tes, daya pembeda, tingkat kesulitan dan tebakan semu.

Parameter peserta tes pada model ini dinyatakan dengan elemen-elemen dari vektor θ_j . Banyaknya elemen dari vektor ini masih merupakan hal yang sering diperdebatkan (Reckase, 1997). Berdasarkan pengalaman Reckase dan Hirsch (Reckase, 1997), banyaknya dimensi kemampuan sering *underestimate* atau *overestimate* dan hal ini akan merugikan. Banyaknya dimensi yang digunakan pada model tergantung interaksi butir dengan peserta tes yang perlu disesuaikan dengan tujuan analisis.

Diskriminasi butir pada teori respons butir multidimensi (*MDISC_i*) merupakan parameter untuk model yang dinyatakan dengan vektor a yang fungsinya mirip dengan parameter a pada teori respons butir unidimensi. Unsur-unsur vektor terkait dengan

kemiringan dari permukaan respons pada arah yang bersesuaian dengan sumbu- m . Kemiringan ini mengindikasikan sensitivitas butir terhadap kemampuan sepanjang sumbu- m . Jika parameter ini mengukur bukan hanya satu dimensi saja, maka diskriminasi butir dapat dinyatakan dengan kombinasi dimensi-dimensi, yang dinyatakan dengan

$$MDISC_i = \sqrt{\sum_{m=1}^k a_{im}^2} \dots\dots\dots(4.4)$$

$MDISC_i$ merupakan diskriminasi dari butir i , k banyaknya dimensi pada ruang- θ , dan a_{im} merupakan elemen dari vektor a untuk butir ke- i .

Tingkat kesulitan butir merupakan parameter d_i pada model. Parameter ini tidak dapat diinterpretasikan dengan cara yang sama dengan parameter- b pada teori respons butir unidimensi. Misalkan a merupakan parameter daya pembeda butir pada model unidimensi, maka $-ab = d_i$. Nilai yang ekuivalen dengan tingkat kesulitan pada model unidimensi dinyatakan dengan

$$MDIFF_i = \frac{-d_i}{MDISC_i} \dots\dots\dots(4.5)$$

$MDIFF_i$ menyatakan jarak dari titik asal ruang- θ terhadap titik kemiringan paling tinggi pada arah dari titik asal. Menurut Ackerman, Gierl, & Walker (2003), tanda dari jarak ini mengindikasikan kesulitan relatif butir. Sebagai contoh, pada tes yang memuat dua dimensi, butir dengan $MDIFF_i$ negatif, relatif mudah dan berada di kuadran III; dan relatif sulit jika terletak di kuadran I. $MDIFF_i$ analog dengan parameter b pada teori respons butir unidimensi. Lokasi (parameter lokasi) bersesuaian dengan arah sudut butir dari tiap butir relatif terhadap sumbu- 1 positif. Arah kemiringan yang paling besar dari titik pusat koordinat, menurut Reckase (1997) dan Ackerman, Gierl, & Walker (2003) dinyatakan dengan

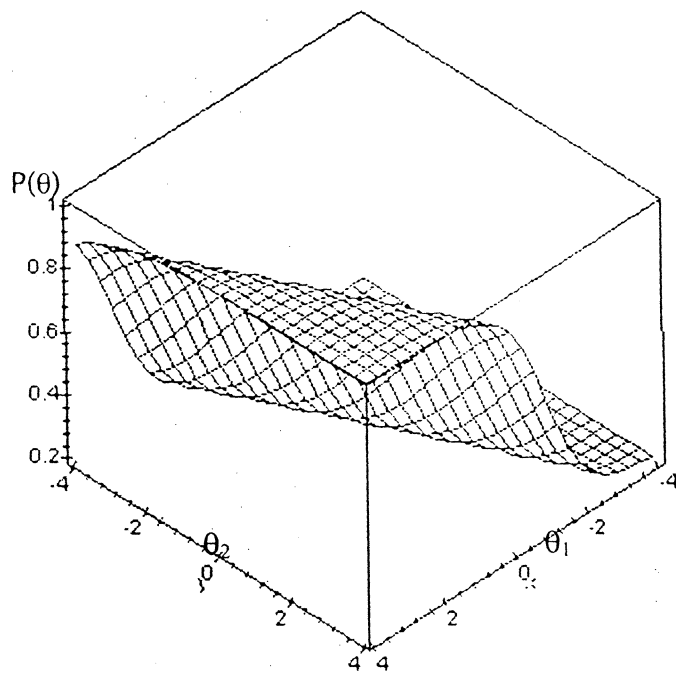
$$\alpha_i = \arccos \frac{a_{im}}{MDISC_i} \dots\dots\dots(4.6)$$

dengan a_{im} merupakan sudut antara garis dari titik pusat koordinat ke titik yang memiliki kemiringan terbesar dengan sumbu ke- m untuk butir ke- i .

Asimtot bawah (*lower asymptote*) merupakan nilai yang menyatakan probabilitas menjawab benar peserta tes ketika kemampuan yang dimilikinya sangat rendah pada

keseluruhan dimensi. Parameter ini sama artinya dengan parameter c pada teori respons butir unidimensi.

Jika dibandingkan dengan model logistik pada teori respons butir unidimensi, perbedaan ini akan sangat mencolok dengan mencermati kurva karakteristik butir pada model logistik multidimensi. Sebagai ilustrasi, pada Gambar 4.1 disajikan kurva karakteristik butir yang memuat dua dimensi, dengan parameter $a_1=1$, $a_2=1$, $d = 1$, $c=0,2$. Kurva karakteristik butir pada model ini nampak sebagai permukaan, sehingga sering disebut pula dengan permukaan respons butir (*Item Response Surface, IRS*) (Bolt & Lall, 2003) atau permukaan karakteristik butir (*Item Characteristic Surface, ICS*) (Ackerman, Gierl, & Walker, 2003). Permukaan respons butir ini akan sangat sulit digambarkan jika dimensi kemampuan yang diukur suatu butir tes lebih dari dua.



Gambar 4.1
Permukaan Karakteristik Butir yang Mengukur Dua dimensi, dengan Parameter $a_1=1$, $a_2=1$,
 $d = 1$, $c=0,2$

Fungsi informasi butir pada teori respons butir multidimensi dinyatakan dengan :

$$I_{i\alpha}(\theta) = \frac{[\nabla_{\alpha} P_i(\theta)]^2}{P_i(\theta)[1 - P_i(\theta)]} \dots\dots\dots(4.7)$$

Dengan $I_{i\alpha}(\theta)$ merupakan informasi yang disajikan oleh butir-i pada arah α dalam ruang dan ∇_{α} merupakan operator definitif untuk turunan dengan arah α . DeBryant (tth) menyajikan fungsi informasi butir berarah ini secara lebih detail. Fungsi informasi berarah disajikan sebagai :

$$I_i(\theta_j) = D^2 (a_i' u_i)^2 Q_i(\theta_j) \{ P_i(\theta_j) [1 + \text{Exp}(-L)]^2 \}^{-1} \dots (4.8)$$

dengan $L = D (a_i' \theta_j + d_i)$. Skala kemampuan yang memaksimumkan nilai fungsi informasi yakni

$$\theta_{\max} = u_i [\ln \{ .5 [1 + (8c_i + 1)^{1/2}] \} (D''a_i''^{-1} - d_i (''a_i''^{-1})) \dots (4.9)$$

Atau jika dinyatakan dengan tingkat kesulitan (MDIFF) dan indeks daya pembeda butir (MDISC) menjadi

$$\theta_{\max} = u_i [\ln \{ .5 [1 + (8c_i + 1)^{1/2}] \} (D.MDISC_i)^{-1} + MDIFF_i] \dots (4.10)$$

Sebagai akibatnya, kemampuan yang memaksimumkan IIF pada dimensi ke-m yakni

$$\theta_{\max m} = [\ln \{ .5 [1 + (8c_i + 1)^{1/2}] \} (D.MDISC_i)^{-1} + MDIFF_i] \cos \alpha_{mi} \dots (4.11)$$

Pada teori respons butir multidimensi, parameter-parameter pada model dapat diestimasi dengan menggunakan berbagai prosedur, misalnya *joint maximum likelihood procedures* (Reckase, 1997). Prosedur estimasi ini bertujuan untuk menemukan himpunan parameter butir dan peserta tes yang akan memaksimumkan likelihood (L) dari respons butir yang teramati. Bentuk persamaan likelihoodnya diberikan oleh:

$$L = \prod_{j=1}^N \prod_{i=1}^n P(u_{ij} | a_i, d_i, c_i, \theta_j) \dots \dots \dots (4.12)$$

dengan u_{ij} merupakan respons butir-i oleh orang j, baik 0 ataupun 1. Terkait dengan perhitungan secara matematis, biasanya program komputer yang ada meminimumkan negatif logaritma dari L atau

$$F = -\ln(L).$$

Pengujian kegunaan model multidimensi yang diusulkan dilakukan menganalisis kecocokan model (*goodness of fit*). Salah satu prosedur yang dapat digunakan adalah cara yang disarankan Reckase (1997). Cara ini ditempuh dengan menguji unsur-unsur matriks korelasi residual antarbutir. Korelasi residual antara butir i_1 dan i_2 , dinyatakan

dengan $r_{i_1 i_2}$ yang dihitung dengan:

$$r_{i_1 i_2} = \frac{1}{n} \sum_{a=1}^N \frac{(u_{i_1 j} - P_{i_1 j})(u_{i_2 j} - P_{i_2 j})}{\sqrt{P_{i_1 j} Q_{i_1 j}} \sqrt{P_{i_2 j} Q_{i_2 j}}} \dots\dots\dots(4.13)$$

dengan u_{ij} merupakan respons peserta butir ke- i dan peserta ke- j , P_{ij} merupakan probabilitas menjawab benar butir ke- i untuk peserta ke- j , dan $Q_{ij} = 1 - P_{ij}$. Kecocokan model yang baik menghasilkan residu estimasi korelasi hasil observasi antarbutir mendekati 0 ketika N besar.

Dalam bidang pendidikan dan psikologi, teori respons butir multidimensi dapat diterapkan untuk mengukur kemampuan umum ataupun kemampuan psikologis tertentu peserta tes, jika tes bersifat multidimensi. Penerapan pendekatan ini terkait dengan banyak hal. Menurut Ackerman, Gierl, & Walker (2003), teori respons butir multidimensi dapat diarahkan pada pengembangan tes, memperoleh informasi diagnostik tentang estimasi kemampuan, keberfungsian butir diferensial (*differential item functioning, DIF*), dan model teori respons butir untuk data politomous. Segall (2000) memperkuat pendapat ini dengan menyatakan bahwa teori respons butir multidimensi dapat digunakan untuk pemilihan butir, dalam rangka memprediksi pembelajaran maupun mengestimasi kemampuan peserta didik.

Sampai saat ini, perangkat lunak (*software*) yang ada untuk menganalisis butir dengan teori respons butir multidimensi (*Multidimensional Item Response Theory, MIRT*) hanyalah untuk model kompensatori. Menurut Spencer (2004) pada data dikotomi, program yang biasa digunakan adalah MAXLOG, NOHARM dan TESTFACT (Spencer, 2004). MAXLOG mengestimasi parameter model kompensatori untuk 2 parameter.

TESTFACT dikembangkan oleh Wilson, Woods, & Gibbons (1984), merupakan program komputer yang dikembangkan untuk menyusun suatu model nonlinear, analisis faktor eksploratori informasi penuh pada respons butir dikotomi. Sebagai program eksploratori, batasan awal untuk parameter butir tidak dispesifikasi. Pada analisis ini, banyaknya kemampuan laten yang dihipotesiskan untuk menjadi perhatian pada konstruksi tes harus dispesifikasikan. Model memprediksi struktur dimensi dari butir individual berdasarkan atas banyaknya kemampuan, didefinisikan secara apriori, yang memberikan kontribusi pada respons-responsnya (McDonald, 1999). Struktur dimensi ini diketahui dengan menggunakan estimasi *marginal maximum likelihood (MML)* yang

dikombinasikan dengan algoritma *expectation-maximization* (EM) yang dikembangkan oleh Bock & Atkin (1981). Algoritma membandingkan peserta tes menjadi sampel acak dari populasi dan mengasumsikan level kemampuan latennya berasal dari populasi yang berdistribusi normal baku (distribusi normal dengan rerata 0 dan standar deviasi 1 (Knol & Berger, 1991). Prosedur ini iteratif, banyaknya harapan dari peserta tes pada setiap level kemampuan dihitung terlebih dahulu dengan banyaknya harapan dari orang yang menjawab dengan setiap butir dengan benar. Kemudian, dengan menggunakan persamaan estimasi MML, berdasarkan model IRT multidimensi ogive normal, estimasi parameter butir dilakukan untuk memaksimalkan likelihood yang diberikan oleh respons butir yang diamati. Parameter butir digunakan untuk mengestimasi ulang frekuensi harapan, yang kemudian ditempatkan sekali lagi dalam persamaan MML dan seterusnya. Sekali frekuensi harapan konvergen (mengumpul) dengan pola respons yang diketahui, parameter butir akhir ditemukan menggunakan prosedur Newton-Gauss (Embretson & Reise, 2000).

Tabel 4.1
Karakteristik Butir Diestimasi dengan Teori Respons Butir 2 Dimensi

No. Butir	Materi	c	b	a1	a2
1	Persentase (soal cerita)	0,017	1,582	1,294	-0,243
2	Diagram Venn	0,018	1,091	1,004	-0,067
3	Persentase	0,045	0,035	0,689	0,082
4	HP bil bulat	0,035	-0,202	0,598	-0,126
5	Jaring-jaring kubus	0,144	2,235	0,801	-0,287
6	Simetri lipat	0,070	1,062	0,536	-0,255
7	Sudut segitiga	0,500	1,023	1,535	-0,093
8	Pemetaan	0,026	1,090	1,037	0,028
9	Akar dan pangkat	0,033	1,444	1,100	-0,201
10	Sifat garis sejajar	0,036	0,151	0,755	-0,106
11	Keliling belah ketupat	0,022	0,963	1,364	-0,017
12	Luas jajar genjang	0,026	0,612	1,183	-0,060
13	Perbandingan (soal cerita)	0,159	2,176	1,039	-0,124
14	Persamaan garis lurus	0,046	0,007	0,813	0,020
15	SPL (soal cerita)	0,041	1,449	1,095	-0,042
16	Median data	0,002	0,366	0,813	0,165
17	Volume limas	0,014	0,175	0,882	0,225
18	Luas permukaan prisma	0,105	0,760	1,107	0,236
19	Refleksi	0,255	0,800	1,154	0,320
20	Dilatasi	0,094	0,623	0,79	0,273

21	Perbandingan segitiga	0,352	-1,761	2,612	0,597
22	Segitiga kongruen	0,038	0,835	0,651	0,042
23	Juring lingkaran	0,359	-0,037	1,098	0,335
24	Persekutuan lingkaran	0,136	0,807	1,099	0,520
25	Suku dan faktor	0,049	0,735	0,874	0,359
26	Fungsi kuadrat	0,116	0,577	1,066	0,414
27	Phytagoras dan luas segitiga	0,292	0,895	1,227	0,438
28	Barisan dan deret	0,035	1,888	1,211	0,169
29	Trigonometri	0,500	0,075	1,353	0,781
30	Logaritma	0,500	-0,219	1,734	0,856

Tabel 4.2
Karakteristik Butir Diestimasi Dengan Teori Respons Butir 3 Dimensi

No. Butir	Materi	c	b	a1	a2	a3
1	Persentase (soal cerita)	0,017	3,89	3,579	0,764	1,638
2	Diagram Venn	0,018	3,395	2,036	0,398	3,802
3	Persentase	0,045	0,11	3,925	0,575	1,842
4	HP bilangan bulat	0,035	-0,469	1,602	0,357	1,701
5	Jaring-jaring kubus	0,144	3,224	1,289	0,215	1,170
6	Simetri lipat	0,070	2,083	1,396	0,333	1,582
7	Sudut segitiga	0,500	2,25	2,665	0,409	1,122
8	Pemetaan	0,026	3,057	1,729	0,492	3,519
9	Akar dan pangkat	0,033	3,284	2,837	0,502	1,718
10	Sifat garis sejajar	0,036	0,416	2,809	0,503	2,046
11	Keliling belah ketupat	0,022	1,872	1,643	0,429	2,503
12	Luas Jajar genjang	0,026	1,764	3,500	0,545	2,319
13	Perbandingan (soal cerita)	0,159	4,799	2,455	0,509	2,107
14	Persamaan garis lurus	0,046	0,013	2,028	0,326	2,184
15	SPL (soal cerita)	0,041	3,274	2,465	0,496	2,117
16	Median data	0,002	1,278	3,638	0,503	2,395
17	Volume limas	0,014	0,571	2,312	0,469	3,283
18	Luas permukaan prisma	0,105	2,292	3,557	0,435	2,492
19	Refleksi	0,255	1,582	2,220	0,363	1,741
20	Dilatasi	0,094	1,687	2,237	0,328	2,548
21	Perbandingan segitiga	0,352	-1,917	2,201	0,284	1,939
22	Segitiga kongruen	0,038	2,748	3,097	0,411	2,291
23	Juring lingkaran	0,359	-0,02	2,729	0,489	3,010
24	Persekutuan lingkaran	0,136	2,166	3,012	0,468	2,704
25	Suku dan faktor	0,049	2,383	3,586	0,193	2,527
26	Fungsi kuadrat	0,116	1,576	2,266	0,716	3,136
27	Phytagoras dan luas segitiga	0,292	1,175	1,574	-0,037	0,921

28	Barisan dan deret	0,035	5,139	0,786	-4,25	0,835
29	Trigonometri	0,500	0,347	0,653	1,583	0,861
30	Logaritma	0,500	-0,208	2,212	3,226	1,907

Sebagai contoh misalnya analisis butir soal Ujian Nasional mata pelajaran matematika SMP (Heri Retnawati, 2008). Setelah melalui analisis faktor untuk membuktikan bahwa ada 2 dan 3 dimensi yang terukur dalam perangkat tes, selanjutnya respons siswa dianalisis menggunakan TESTFACT. Hasil yang diperoleh berupa parameter butir dan parameter kemampuan untuk tiap dimensi. Hasil estimasi disajikan pada Tabel 4.1 untuk model 2 dimensi dan Tabel 4.2 untuk model 3 dimensi.

Pada kasus analisis butir dengan menggunakan teori respons butir multidimensi pada data contoh data tersebut, salah satu output yang dihasilkan yakni *factor score* yang menunjukkan kemampuan peserta tes (*ability*) untuk tiap dimensi. Kemampuan peserta ini beradapada skala kemampuan $[-3,3]$ yang kemudian dapat disajikan ke skala 100 melalui transformasi linear. Hasil analisis statistik deskriptif untuk tiap dimensi disajikan pada Tabel 4.3.

Tabel 4.3

Statistik Deskriptif Kemampuan Peserta

Statistik	Dimensi					
	1		2		3	
	Umum	Umum	Spasial	Umum	Spasial	Numerik
Rerata	49,2180	49,4526	50,1869	49,6449	50,3513	49,8990
SD	16,3195	16,3311	9,2513	16,1535	8,8616	8,8271
Minimm	1,8570	1,4903	15,0430	3,0406	24,6283	20,4108
Maksimum	77,4055	77,6889	83,3400	77,7222	87,7909	74,6883

Membandingkan ketiga hasil analisis menggunakan model 1, 2, dan 3 dimensi tersebut, model analisis yang paling teliti yakni model 3 dimensi. Hal ini dapat dimengerti karena analisis dengan model lebih dari satu dimensi, akan diperoleh informasi yang lebih detail tentang kemampuan peserta. Pada penelitian ini, jika hanya 1 dimensi saja kemampuan yang diukur, maka informasi yang diperoleh hanya informasi umum tentang kemampuan matematika umum saja. Dengan analisis dua dimensi, akan dapat diketahui kemampuan

umum dan spasial. Jika dianalisis dengan model 3 dimensi, ada 3 kemampuan yang dapat terukur, yakni kemampuan umum, kemampuan spasial, dan kemampuan numerik. Hasil ini dapat dipahami, butir-butir tidak hanya mengukur kemampuan matematika umum saja, namun juga mengukur kemampuan yang lainnya. Sebagai contoh butir nomor 14 dan butir 16 pada perangkat UN mata pelajaran matematika tahun 2006.

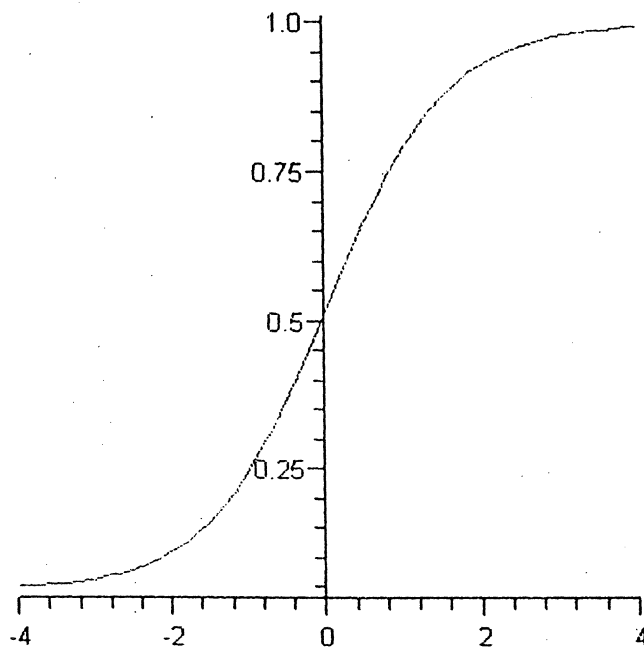
Naskah butir nomor 14 sebagai berikut.

14. Persamaan garis yang melalui titik (1,-3) dan tegak lurus terhadap garis dengan persamaan $y = \frac{2}{3}x + 5$ adalah
- a. $2x - 3y - 3 = 0$
 - b. $2x + 3y + 3 = 0$
 - c. $3x + 2y + 3 = 0$
 - d. $3x + 2y - 3 = 0$

Kunci jawaban : D

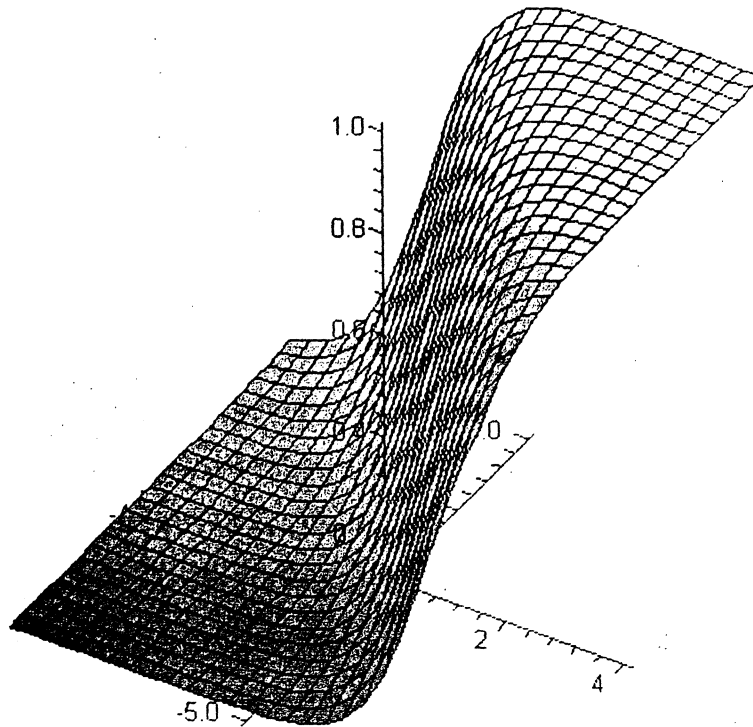
Untuk dapat menjawab benar butir ini, peserta tes perlu mengetahui konsep garis yang tegak lurus dengan garis lain dan kondisinya. Setelah mengetahui bahwa garis yang tegak lurus dengan garis $y = \frac{2}{3}x + 5$ memiliki gradien $-\frac{3}{2}$, barulah membentuk persamaan garis dengan gradien $-\frac{3}{2}$ yang kemudian diselesaikan. Untuk mengerjakan butir ini, paling tidak ada dua kemampuan yang diperlukan. Kemampuan yang pertama terkait dengan kemampuan umum yakni memahami suatu persamaan linear dan menyelesaikannya. Kemampuan kedua, mengenai konsep suatu garis yang tegak lurus dengan garis lain.

Dianalisis dengan teori respons butir unidimensi, butir nomor 14 tersebut memiliki parameter tingkat kesulitan sebesar 0,018, daya pembeda sebesar 0,762 dan parameter *pseudo guessing* sebesar 0,046. Parameter tingkat kesulitan ini termasuk kategori sedang, sehingga butir ini bukan merupakan butir yang sulit. Setelah diketahui parameternya, kurva karakteristik dapat disajikan pada Gambar 4.4.



Gambar 4.4
Kurva Karakteristik Butir nomor 14 Perangkat UN
Mata Pelajaran Matematika 2006

Analisis butir 14 dengan pendekatan teori respons butir bidimensi menghasilkan parameter tingkat kesulitan sebesar 0,007, parameter *pseudo guessing* sebesar 0,046 dan daya pembeda untuk dimensi kemampuan umum sebesar 0,813 dan dimensi spasial sebesar 0,020. Permukaan karakteristik butir disajikan pada Gambar 4.5. Pada analisis dengan pendekatan 3 dimensi, di hasilkan parameter tingkat kesulitan sebesar 0,013, parameter *pseudo guessing* sebesar 0,046. dan daya pembeda untuk dimensi kemampuan umum sebesar 2,028, dimensi spasial sebesar 0,326, dan kemampuan numerik sebesar 2,184. Namun pada model 3 dimensi, permukaan karakteristik butir tidak dapat digambarkan lagi.



Gambar 4.5
Permukaan Karakteristik Butir Nomor 14 Perangkat UN
Mata Pelajaran Matematika 2006

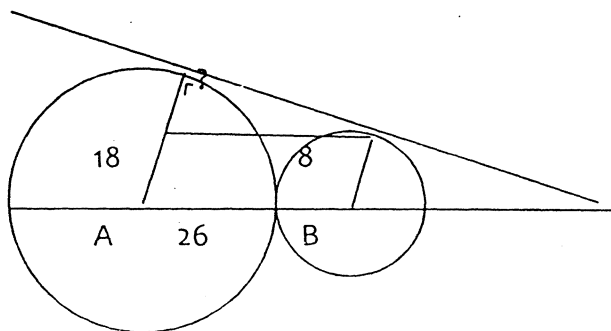
Mencermati Gambar 4.4 pada kurva karakteristik butir (unidimensi), dibandingkan dengan permukaan karakteristik butir (bidimensi) pada Gambar 4.5, diperoleh bahwa model bidimensi lebih teliti. Pada model unidimensi, hanya ada 1 informasi kemampuan yang diperoleh yakni kemampuan matematika umum (θ) yang dapat diperoleh. Pada model bidimensi, ada 2 kemampuan yang dapat terukur, yakni kemampuan umum (θ_1) dan kemampuan spasial (θ_2).

Naskah butir nomor 24 perangkat UN mata pelajaran matematika 2006 sebagai berikut.

24. Dua lingkaran A dan B masing-masing berdiameter 36 cm dan 16 cm. Jika jarak AB = 26 cm, panjang garis singgung persekutuan luar kedua lingkaran tersebut adalah
- 22 cm
 - 24 cm
 - 26 cm
 - 28 cm

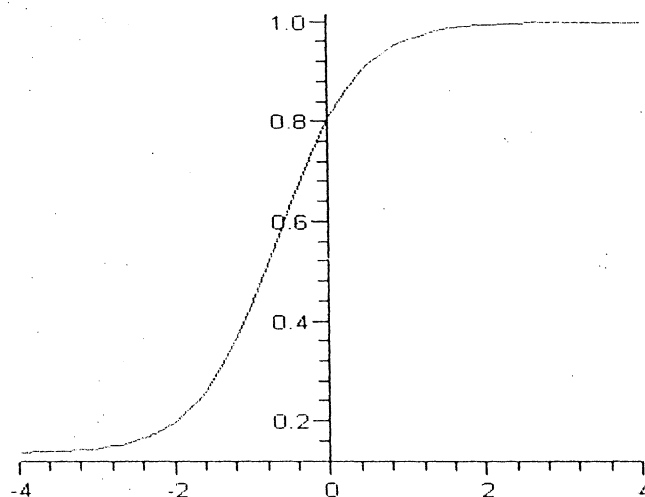
Kunci: B

Untuk dapat menjawab benar butir ini, peserta perlu terlebih dahulu mengetahui bahwa jarak AB sama dengan jumlah jari-jari kedua lingkaran. Sebagai akibatnya, kedua lingkaran ini bersekutu di satu titik. Peserta tes perlu membuat sketsa, kemudian meletakkan jarak AB dan panjang jari-jari lingkaran. Dengan menggunakan sifat garis singgung lingkaran berpotongan secara tegak lurus dengan jari-jari lingkaran pada titik singgung dan teorema Pythagoras, dapat dihitung panjang garis singgung persekutuan luar kedua lingkaran tersebut, yakni 24 cm. Sketsa untuk menyelesaikan butir nomor 24 disajikan pada Gambar 4.6.



Gambar 4.6

Sketsa untuk Menyelesaikan Butir Nomor 24 Perangkat UN
Mata Pelajaran Matematika Tahun 2006

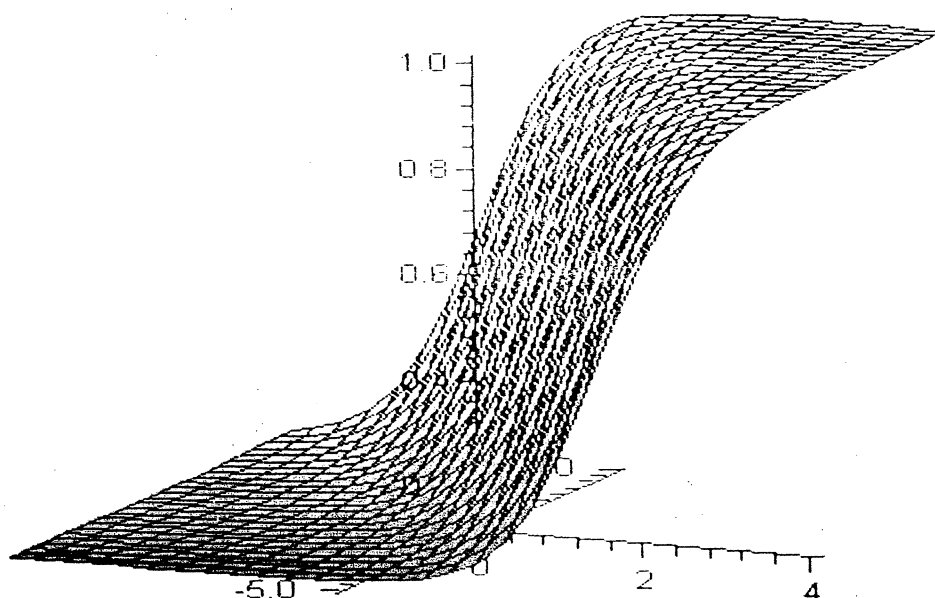


Gambar 4.7

Kurva Karakteristik Butir Nomor 24 Perangkat UN
Mata Pelajaran Matematika 2006

Jika dianalisis dengan teori respons butir unidimensi, butir nomor 24 tersebut memiliki parameter tingkat kesulitan sebesar $-0,667$, daya pembeda sebesar $1,114$ dan parameter *pseudo guessing* sebesar $0,136$. Parameter tingkat kesulitan ini termasuk kategori sedang, sehingga butir ini bukan merupakan butir yang mudah ataupun butir yang sulit. Kurva karakteristik dapat disajikan pada Gambar 4.7.

Analisis butir 24 dengan pendekatan teori respons butir bidimensi menghasilkan parameter tingkat kesulitan sebesar $0,807$, parameter *pseudo guessing* sebesar $0,136$ dan daya pembeda untuk dimensi kemampuan umum sebesar $1,099$ dan dimensi spasial sebesar $0,520$. Permukaan karakteristik butir disajikan pada Gambar 4.8. Pada analisis dengan pendekatan 3 dimensi, dihasilkan parameter tingkat kesulitan sebesar $2,166$, parameter *pseudo guessing* sebesar $0,136$. dan daya pembeda untuk dimensi kemampuan umum sebesar $3,012$, dimensi spasial sebesar $0,468$, dan dimensi numerik sebesar $2,704$. Permukaan karakteristik butir model 3 dimensi juga tidak dapat digambarkan lagi.



Gambar 4.8
Permukaan Karakteristik Butir Nomor 24 Perangkat UN
Mata Pelajaran Matematika 2006

* * *

Daftar Pustaka

- Ackerman, T.A., Gierl, M.J., & Walker, C.M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement*, Vol. 22, pp. 37-53.
- Bock, R.D. & Atkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algoritm. *Psychometrica*, No. 46, pp. 443-459.
- Bolt, D.M. & Lall, V.M. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Marcov chain Monte-Carlo. *Applied Psychological Measurement*, No. 27, pp. 395-414.
- De Bryant, U. (tth). Directional item information for the multidimensional three-parameter logistik model. *Running head: Multidimensional item information*. Diambil dari <http://pegasus.cc/ucf/edu/> pada tanggal 2 November 2006.
- Embretson, S.E. & Reise, S.P., (2000). *Item response theory for psychologists*. Marwah, NJ: Lawrence Erlbaum.
- Heri Retnawati (2008). Estimasi efisiensi relative tes berdasarkan teori tes klasik dan teori respons butir. *Disertasi*. Universitas Negeri Yogyakarta, tidak dipublikasikan.
- Knol, D.L. & Berger, MP.F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, No. 26, pp. 457-477.
- Reckase, M.D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W.J. Linden & R.K. Hambleton (Eds), *Handbook of modern item response theory* (pp. 271-286). New York: Springer.
- Segall, D.O. (2000). General ability measurement: An application of multidimensional item response theory. *Psychometrica*, Vol. 66, 79-97.
- Spencer, S.G. (2004). The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated. *Dissertation*. Brigham Young University.
- Spray, J.A., Davey, T.C., Rechase, M.D., et al. (1990). Comparison of two logistic multidimensional item response theory models. *ACT Research Report Series*. United States Government.
- Wilson, D., Wood, R. & Gibbons, R. (1984). *TESTFACT: Test scoring and fullinformation item factor analysis*. [Computer program]. Mooresville, IL: SSI.

BAB 5

PENGEMBANGAN BANK SOAL



Evaluasi dalam pendidikan dilaksanakan untuk memperoleh informasi tentang aspek yang berkaitan dengan pendidikan. Menurut Gronlund (1976: 8), evaluasi dalam pendidikan memiliki tujuan : a) untuk memberikan klarifikasi tentang sifat hasil pembelajaran yang telah dilaksanakan, b) memberikan informasi tentang ketercapaian tujuan jangka pendek yang telah dilaksanakan, c) memberikan masukan untuk kemajuan pembelajaran, d) memberikan informasi tentang kesulitan dalam pembelajaran dan untuk memilih pengalaman pembelajaran di masa yang akan datang. Informasi evaluasi dapat digunakan untuk membantu memutuskan a) kesesuaian dan keberlangsungan dari tujuan pembelajaran, b) kegunaan materi pembelajaran, dan c) untuk mengetahui tingkat efisiensi dan efektifitas dari strategi pengajaran (metode dan teknik belajar-mengajar) yang digunakan.

Evaluasi memiliki fungsi untuk membantu guru dalam hal-hal : a) penempatan siswa dalam kelompok-kelompok tertentu, b) perbaikan metode mengajar, c) mengetahui kesiapan siswa (sikap, mental, material), d) memberikan bimbingan dan seleksi dalam rangka menentukan jenis jurusan maupun kenaikan tingkat (Gronlund, 1976: 16). Dalam evaluasi pendidikan, diperlukan alat (instrumen). Alat yang digunakan untuk melakukan evaluasi, salah satunya adalah tes. Tes ini digunakan untuk mengetahui informasi tentang aspek psikologis tertentu. Menurut Cronbach (1970), tes merupakan suatu prosedur sistematis untuk mengamati dan menggambarkan satu atau lebih karakteristik seseorang dengan suatu skala numerik atau sistem kategorik. Berdasarkan hal ini, tes memberikan informasi yang bersifat kualitatif dan kuantitatif.

Tes dapat diklasifikasikan dengan beberapa macam, tergantung dari tujuannya (Anastasi dan Urbina, 1997 : 2-4). Tes prestasi belajar merupakan suatu bentuk tes untuk

mendapatkan data, yang merupakan informasi untuk melihat seberapa banyak pengetahuan yang telah dimiliki dan dikuasai oleh seseorang sebagai akibat dari pendidikan dan pelatihan (Anastasi dan Urbina, 1997: 42-43). Berdasarkan informasi yang diperoleh ini, pada proses seleksi, siswa dapat dikelompokkan sesuai dengan kemampuannya, yang diterima atau tidak diterima. Hal ini sesuai dengan fungsi tes prestasi seperti yang dikemukakan Gronlund (1976: 16), yang menyatakan bahwa tes prestasi berfungsi sebagai alat untuk penempatan, fungsi formatif, fungsi diagnostik dan fungsi sumatif.

Berdasarkan bentuknya, tes prestasi belajar dapat dikelompokkan menjadi tiga jenis, yaitu : 1) objektif, yang sederhana terdiri dari bentuk jawaban singkat, benar-salah atau dua pilihan, dan menjodohkan, serta objektif pilihan ganda dengan alternatif jawaban lebih dari dua, 2) uraian. Hal ini sesuai dengan yang dinyatakan Gronlund (1976: 144) sebagai berikut.

The items used in classroom tests are typically divided into two general categories: (1) the objective item which is highly structured and requires the pupil to supply a word or two or to select the correct answer from among a limited number of alternatives, and (2) the essay question which permits the pupil to select, organize, and present his essay form.

Demikian pula halnya dengan tes dalam pendidikan matematika. Untuk dapat mengetahui kemampuan matematika siswa, baik kemampuan awal maupun hasil belajar, diperlukan suatu evaluasi. Salah satu bentuknya adalah tes. Agar tes yang dilakukan dapat mengetahui kemampuan matematika siswa yang sebenarnya, diperlukan suatu perangkat tes yang baik.

Perangkat tes kemampuan matematika yang baik dapat ditinjau dari berbagai sisi. Pertama, isi tes sebaiknya sesuai dengan materi yang hendak diujikan, sehingga validitasnya baik. Kedua, tes memiliki konstruk yang baik. Ketiga, tes yang baik harus memiliki keajegan (*reliable*). Jika digunakan untuk mengukur beberapa kali, baik pada peserta tes yang sama ataupun berbeda, hasilnya relatif sama.

Suatu perangkat tes yang baik tersusun atas butir-butir soal yang baik. Butir-butir soal yang baik yang digunakan pada perakitan perangkat tes dapat diperoleh dari bank soal. Dalam bank soal, karakteristik butir-butir penyusunnya dapat diketahui karakteristiknya.

Pengertian Bank Soal

Secara singkat, bank soal yang biasa dikenal pendidik didefinisikan sebagai kumpulan dari butir-butir tes. Namun bank soal tidak hanya mengacu pada sekumpulan soal-soal saja. Bank soal mengacu pada proses pengumpulan soal-soal, pemantauan dan penyimpanannya dengan informasi yang terkait sehingga mempermudah pengambilannya untuk merakit soal-soal (Thorndike, 1982). Millman (dalam J. Umar, 1999) mendefinisikan bank soal sebagai kumpulan yang relative besar, yang mempermudah dalam memperoleh pertanyaan-pertanyaan penyusun tes. “Mudah” memiliki pengertian bahwa soal-soal tersebut diberi indeks, terstruktur, dan diberi keterangan sehingga mudah dalam pemilihannya untuk disusun sebagai perangkat tes pada suatu ujian.

Senada dengan pengertian-pengertian di atas, Choppin (dalam J. Umar, 1999) memberikan definisi bahwa bank soal merupakan sekumpulan dari butir-butir tes yang diorganisasikan dan dikatalogan untuk mencapai jumlah tertentu berdasarkan isi dan juga karakteristik butir. Karakteristik butir ini meliputi tingkat kesulitan, reliabilitas, validitas dan lain-lain.

Dari definisi beberapa ahli, sebagian besar mengharuskan penyimpanan bank soal di dalam computer. Dalam pengembangan bank soal kecil, memang mungkin dilakukan tanpa bantuan computer. Tetapi dalam pengembangan bank soal yang besar, tidak mungkin mengembangkan bank soal tanpa bantuan computer. Hal ini disebabkan karena dalam pengembangan bank soal yang besar, ada beberapa tahapan yang tidak mungkin dilakukan tanpa bantuan computer.

Perlunya Pengembangan Bank Soal

Ide pengembangan bank soal terkait dengan kebutuhan merakit tes lebih mudah, cepat dan efisien. Selain itu juga adanya tuntutan kualitas butir soal yang menyusun tes. Dengan adanya bank soal, kualitas butir-butir soal penyusun tes dapat dijamin kualitasnya. Van der Linden (dalam J. Umar, 1999) menyatakan bahwa pengembangan bank soal merupakan praktek baru dalam pengembangan tes, sebagai hasil dari pengenalan teori respons butir dan kegunaan ekstensif dari pengetahuan computer di masyarakat yang modern.

Pada suatu bank soal yang dikembangkan dengan teori respons butir, program tes dapat dibuat lebih fleksibel dan sesuai. Hal ini disebabkan karena karakteristik butir perangkat tes pada teori respons butir tidak tergantung pada karakteristik peserta tes pada saat kaliberasi. Selain itu, kemampuan siswa peserta

tes dapat diketahui dan dapat dibandingkan, karena parameter kemampuan dapat diestimasi pada skala yang sama (Jahja Umar, 1999).. Terkait dengan perkembangan ilmu dan teknologi, pengembangan bank soal berdasarkan teori respons butir dapat diset untuk dikembangkan menjadi *computerized adaptive testing* (Hambleton, Swaminathan, dan Rogers, 1991).

Keuntungan-keuntungan yang dapat diperoleh dengan adanya pengembangan bank soal sebagai berikut :

- 1) kebijakan desentralisasi pada program tes nasional dapat dikenalkan tanpa mengorbankan dapat dibandingkannya hasil tes,
- 2) biaya dan waktu yang diperlukan pada kegiatan konstruksi tes dapat direduksi,
- 3) semakin besar jumlah butir soal yang terdapat pada bank soal, permasalahan keamanan menjadi lebih terjamin.
- 4) Kualitas program tes dapat ditingkatkan, dengan adanya butir-butir dalam bank soal yang telah diketahui karakteristiknya.
- 5) Pendidik dapat mendesain perangkat tes yang akan digunakannya, dengan memanfaatkan butir-butir yang baik dalam bank soal.
- 6) Guru dapat mengkonsentrasikan diri pada usaha untuk meningkatkan kualitas pembelajaran, tanpa harus membelanjakan waktu banyak untuk penyusunan perangkat tes (Jahja Umar, 1999).

Choppin (dalam Jahja Umar, 1999) berpendapat bahwa keuntungan dalam pengembangan bank soal dapat dikelompokkan menjadi empat kategori, Pertama, kategori ekonomi. Dengan adanya system bank soal, memungkinkan adanya penggunaan butir-butir soal yang baik secara berulang. Kedua, dengan adanya bank soal, panjang tes dapat disesuaikan dengan kebutuhannya, yang merupakan kategori fleksibilitas. Ketiga, kategori konsistensi. Dengan adanya bank soal, dapat dikembangkan tes yang parallel, dan hasilnya pun dapat diperbandingkan karena kemampuan peserta tes dapat diketahui dengan skala yang sama. Kategori keempat keamanan. Dengan adanya bank soal, pengembang tes dapat menyusun beberapa tes alternatif untuk menjaga kebocoran soal pada tes yang tujuannya sangat penting.

Pengembangan Bank Soal

Ada beberapa kegiatan penting dalam pengembangan bank soal. Kegiatan tersebut yakni penulisan butir soal, validasi dan kaliberasi butir soal, penyimpanan

dan pengamanan soal, pengaitannya dengan butir-butir baru dalam bank soal, dan mempertahankan bank soal (Jahja Umar, 1999).

Proses penulisan butir soal merupakan hal yang penting dalam pengembangan bank soal. Penulisan butir soal ini bukan merupakan suatu hal yang mudah. Pada penulisan butir soal, diperlukan rekrutmen dan training bagi penulisnya, yang memerlukan biaya yang besar.

Pada pengembangan bank soal matematika, pada penulisan butir soal ini terlebih dahulu dilihat tujuan tes yang akan dikembangkan menggunakan butir dari bank soal. Apakah tes yang akan dikembangkan tersebut untuk seleksi, tes penalaran, atautkah tes prestasi belajar. Tujuan pengembangan tes perlu diperhatikan mengingat sifat-sifat tes tersebut berbeda-beda.

Hal lain yang perlu diperhatikan pada penulisan butir soal untuk pengembangan bank soal matematika adalah lingkup materi matematika. Dengan memperhatikan lingkup atau cakupan materi yang merupakan bahan tes, diharapkan butir soalnya tidak terlalu mudah atau tidak terlalu sulit. Butir soal seperti ini yang dapat membedakan peserta tes berdasarkan kemampuan matematikanya. Terkait dengan hal ini, pembuatan kisi-kisi terlebih dahulu akan memudahkan penulisan butir soal.

Langkah selanjutnya adalah validasi dan kaliberasi. Pada tahap ini, terlebih dahulu butir-butir soal yang ada disusun menjadi perangkat tes kemudian diujicobakan. Ujicoba disesuaikan dengan peserta tes yang akan merespons perangkat tes. Pada pengembangan bank soal berdasarkan teori tes klasik, peserta ujicoba harus berasal dari berbagai strata siswa secara proporsional. Hal ini disebabkan pada teori tes klasik, karakteristik peserta ujicoba mempengaruhi karakteristik butir soal yang diujicobakan. Jika menggunakan pendekatan teori respons butir, yang perlu diperhatikan adalah jumlah peserta ujicoba, mengingat model parameter berbeda akan memerlukan ukuran peserta ujicoba yang berbeda pula agar karakteristik butirnya stabil (Hambleton dan Swaminathan, 1985).

Validasi merupakan proses menentukan validitas perangkat tes. Validitas ini dapat diketahui dari isi, konstruk, maupun dikorelasikan dengan criteria lainnya. Adapun kaliberasi merupakan proses untuk menentukan karakteristik butir soal. Pada pengembangan bank soal berdasarkan teori tes klasik, diestimasi tingkat kesulitan, daya pembeda dan reliabilitas. Pada teori respons butir diestimasi parameter butirnya. Pada model satu parameter, diestimasi tingkat kesulitannya, estimasi nilai fungsi informasi dan estimasi kesalahan pengukurannya. Pada model dua parameter diestimasi tingkat kesulitan, daya pembedanya, estimasi nilai fungsi

informasi dan estimasi kesalahan pengukurannya, sedang pada model tiga parameter diestimasi tingkat kesulitan, daya pembeda, tebakan semu, estimasi nilai fungsi informasi dan estimasi kesalahan pengukurannya. Agar lebih mudah dilakukan, kaliberasi ini dapat dilakukan dengan bantuan komputer, dengan program Itean, Ascal, Rascal, Bigstep, Bilog, Multilog dan lain-lain.

Dari hasil kaliberasi, dapat ditentukan butir-butir soal yang baik. Butir soal yang baik ini merupakan bank soal yang terjadi. Penyimpanan dan pengamanan butir soal yang terjadi ini merupakan hal yang penting, yang merupakan langkah lanjut dari kaliberasi.

Langkah selanjutnya adalah mengaitkan butir-butir soal yang ada dengan butir soal yang baru (*linking new items*). Langkah ini bertujuan agar butir-butir baru yang ditambahkan dalam bank soal terkait dengan butir-butir yang lama berdasarkan kaliberasi yang telah dilakukan. Prosesnya dinamai dengan penyetaraan (*equiting*), yang bertujuan untuk memastikan kualitas butir soal dan mengestimasi konstanta hubungan dengan perangkat tes yang lama.

Untuk mempertahankan keberadaan bank soal, perlu dilakukan ujicoba ulang dan penambahan butir-butir soal yang baru. Sejarah butir soal hendaknya juga dicatat. Hal ini dilaksanakan untuk menjamin kualitas butir-butir dalam bank soal.

Permasalahan dalam Pengembangan Bank Soal

Ada beberapa permasalahan yang terkait dengan pengembangan bank soal. Berikut ini merupakan permasalahan yang timbul dalam praktek pengembangan bank soal.

- 1) Pengembangan bank soal merupakan investasi yang sangat mahal.
- 2) Pengembangan bank soal memerlukan ahli khusus.
- 3) Konstruksi butir yang memenuhi teori respons butir sangat sulit.
- 4) Pada butir-butir tes prestasi, tuntutan syarat pada teori respons butir sulit untuk dipenuhi (Jahja Umar, 1999).

Terlepas dari pendefinisian bank soal oleh para ahli, pendidik dan pengembang tes matematika dapat memanfaatkan kumpulan butir-butir tes dari bank soal untuk mengevaluasi dengan berbagai tujuan dalam pendidikan matematika. Dengan adanya bank soal matematika, ada jaminan fleksibilitas, efisiensi, kualitas butir perangkat tes, keamanan tes, dan konsistensi pada pelaksanaan tes. Adapun langkah-langkah dalam pengembangan bank soal matematika adalah menulis butir tes matematika, melakukan validasi dan

kaliberasi, penyimpanan dan pengemasan, mengaitkan butir baru dengan butir dalam bank soal dan pemeliharaan bank soal.

Contoh Pengembangan Bank Soal

Contoh bank soal yang dituliskan dalam buku sini dikembangkan oleh Heri Retnawati dan Samsul Hadi (2012-2013), yakni Pengembangan bank soal untuk ujian kenaikan kelas di DI Yogyakarta. Untuk merumuskan model bank soal yang diharapkan, dilakukan *focus group discussion* (FGD). Berdasarkan hasil FGD, dapat diperoleh kesimpulan bahwa selama ini antar kabupaten di DI Yogyakarta pelaksanaan ujian sendiri-sendiri, bahkan sekolah menyusun soalnya sendiri-sendiri. Perangkat tes yang di gunakan antar kabupaten yang satu dengan yang lain merupakan perangkat yang berbeda. Antar perangkat tes yang digunakan tidak ada butir bersama. Namun, keberadaan butir bersama disepakati untuk dibuat bersama dan digunakan bersama oleh peserta FGD agar penskoran berada pada skala yang sama.

Untuk di kabupaten Gunungkidul, pelaksanaan Ujian Kenaikan Kelas (UKK) sebenarnya tanggungjawab sekolah masing-masing, karena setiap guru dan sekolah mempunyai hak untuk menguji, dan penilaian juga perlu dilakukan oleh guru. Sebenarnya yang mempunyai tugas melakukan evaluasi adalah guru, terlebih lagi di era otonomi daerah. Dinas pendidikan pada dasarnya memberikan layanan kepada masyarakat, bentuk salah satunya dalam bentuk penyelenggaraan ujian, termasuk menyediakan perangkat tesnya.

Bank soal di kedua kabupaten belum ada. Selama ini, guru-guru mengembangkan tes dimulai dengan menyusun kisi-kisi yang sesuai dengan indikator dari standar kompetensi dan kompetensi dasar yang akan dicapai pembelajaran. Soal-soal sudah digunakan tidak dimanfaatkan lagi, meskipun guru-guru sudah melakukan analisis butir dan dapat memanfaatkannya untuk perbaikan pembelajaran. Bagi dinas pendidikan, adanya bank soal dan pengembangannya sangat diperlukan dan memudahkan guru merakit soal, dan soal-soalnya pun telah dapat diketahui karakteristiknya. Dengan diketahuinya karakteristik siswa, perangkat soal yang digunakan pada ujian dapat mengukur kemampuan siswa.

Koordinasi antar kabupaten terkait dengan butir bersama belum ada. Koordinasi antar kabupaten baru terkait dengan kalender pendidikan yang difasilitasi oleh dinas Pendidikan Provinsi. Terkait dengan pemanfaatan ke depan, guru-guru di Gunugkidul sangat menyetujui adanya butir bersama, sehingga penskalaan kemampuan menjadi lebih valid. Hal ini juga diperkuat oleh pejabat dinas pendidikan bahwa butir bersama merupakan suatu hal yang diperlukan, agar skala kemampuan berada pada skala yang sama. Dengan adanya skala yang sama, terjadi keadilan ketika melakukan perbandingan kualitas. Pemanfaatan butir bersama juga disarankan yakni untuk pengembangan bank soal.

Menurut pakar pendidikan, di Indonesia, otonomi sampai di tingkat kabupaten, namun sumber daya manusia belum mendukung. Jika seandainya bank soal ada, factor keamanan harus dipikirkan/dipertimbangkan. Sistem dalam bank soal juga perlu dirancang agar memudahkan guru memanfaatkannya. Karena dari beberapa pengalaman, guru merasa kesulitan untuk menyelesaikan masalah-masalah pendidikan, termasuk diantaranya melaksanakan penilaian dan pemanfaatannya. Perlu pula dilaksanakan upaya untuk meningkatkan profesionalisme guru, diantaranya kerjasama guru dan dinas mengembangkan bank soal.

Pakar pengukuran memberikan masukan, bahwa bank soal bukanlah sekumpulan butir. Bank soal lebih ke sistemnya, termasuk menyimpan butir, menambah butir, menghapus butir, menyimpan riwayat butir mulai pembuat, karakteristik dan penggunaannya. Jika bisa penyimpanannya di jaringan sehingga bisa diakses oleh banyak guru. Perlu menjadi perhatian yakni pengamanannya, guru-guru yang menggunakan perlu diberikan username dan password ketika akan mengakses bank soal, sehingga guru MGMP lebih mudah menggunakan, menambah butir, melakukan penghapusan butir, dan lain-lain.

Format bank soal yang biasa digunakan guru dan yang diinginkan oleh guru pada bank soal disajikan pada Gambar 5.1. Format tersebut memuat narasi butir dan identitas butir, baik standar kompetensi, kompetensi dasar, dan indikator soal. Soal-soal ini ditulis atau dicetak manual, kemudian secara manual pula dipindahkan ke format soal ujian.

KARTU SOAL		
Jenis Sekolah Mata Pelajaran Bahan Kuis Bentuk Tes		Nama Penyusun Tahun Pelajaran
Kompetensi Dasar	Buku Sumber	
Materi	No Soal 1	Kuantitas Bank Soal
Indikator Soal	Kunci Jawaban	
(Empty space for content)	(Empty space for content)	

Gambar 5.1. Format Bank Soal

Departemen Pendidikan Nasional melalui sosialisasi KTSP dari pusat kurikulum juga mengeluarkan format bank soal. Meskipun format ini masih manual, namun format ini lebih lengkap karena memuat karakteristik butir. Format bank soal berdasarkan sosialisasi KTSP disajikan pada Gambar 5.2.

KARTU SOAL URAIAN/PRAKTIK

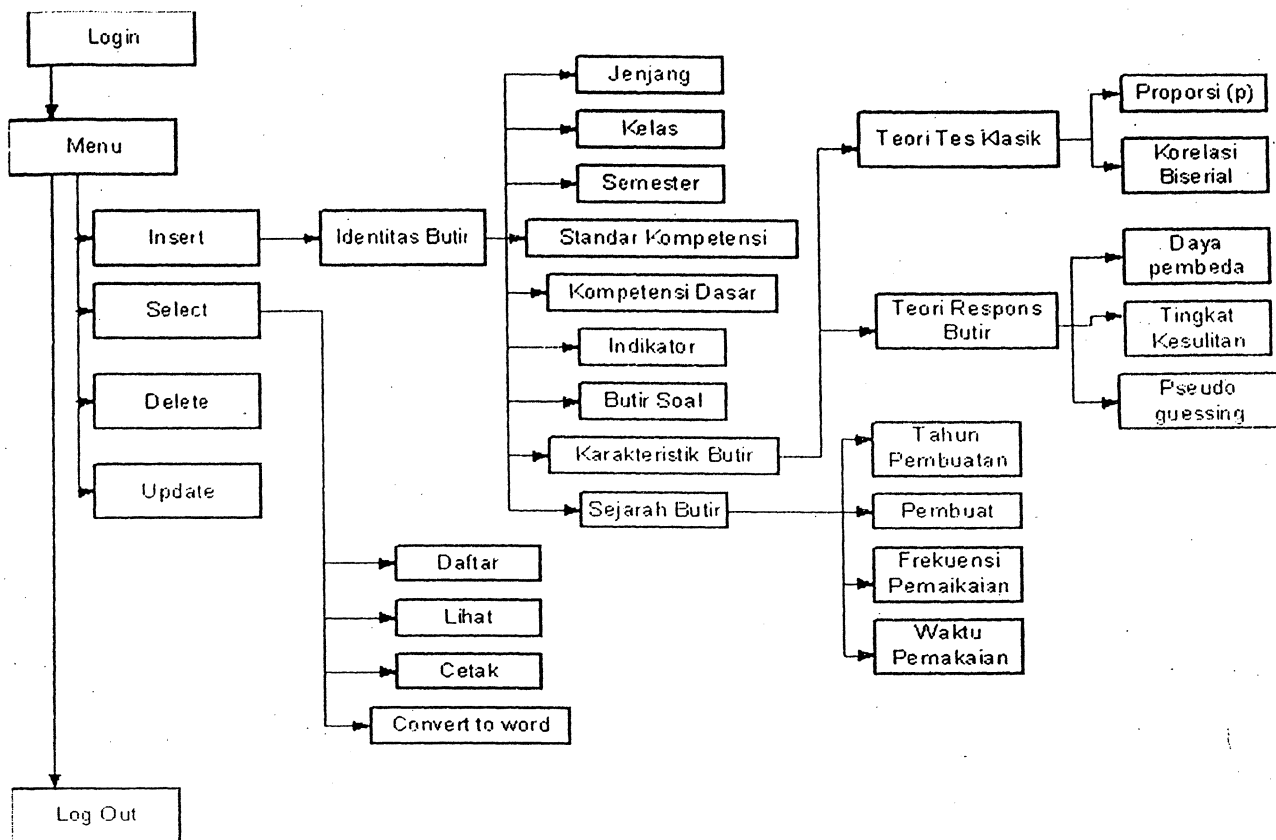
Jenis Sekolah : _____ Penyusun : 1. _____
 Mata Pelajaran : _____ 2. _____
 Bahan Kelas/smt : _____ Tahun ajaran : _____
 Bentuk Tes : Tertulis (Uraian) / Praktik (Kinerja, penugasan, hasil karya)

KOMPETENSI DASAR	NO. SOAL	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">BUKU SUMBER:</td> </tr> </table>	BUKU SUMBER:
BUKU SUMBER:			
MATERI	RUMUSAN BUTIR SOAL		
INDIKATOR SOAL			

No	Digunakan untuk	Tanggal	Jumlah butir	Tingkat kesukaran	Daya pembeda	Proporsi jawaban pada aspek												Keterangan								
						A			B			C			D											
						1	2	3	1	2	3	1	2	3	1	2	3									

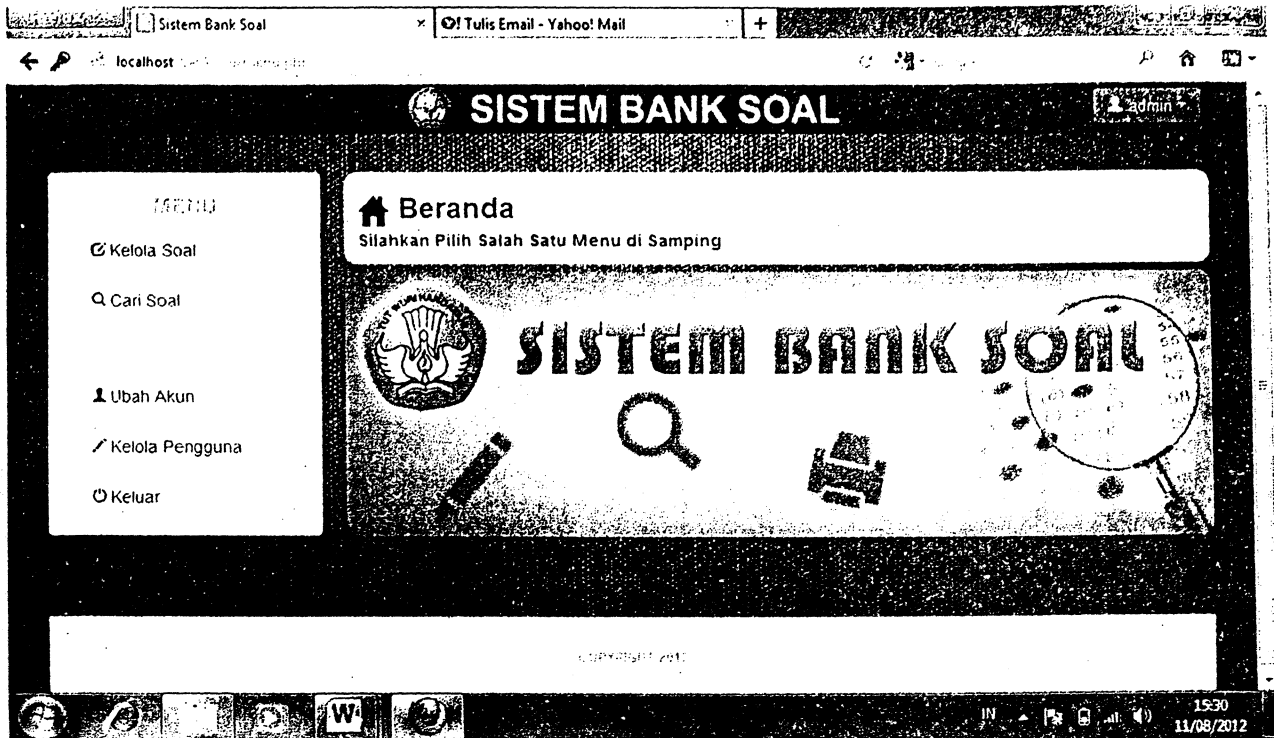
Gambar 5.2. Format Bank Soal dari Pusat Kurikulum

Berdasarkan hasil FGD dan kajian pustaka, model bank soal yang diharapkan yakni sistem yang meliputi penyimpanan butir, pemanfaatan butir, meng-update butir, menghapus butir. Sistem ini dikelola berbasis teknologi informasi dalam satu jaringan yang menyajikan menu-menu. Penyimpanan butir memuat identitas, isi, dan karakteristik butir. Menu butir meliputi *insert*, *select*, *delete*, dan *update*. Pada pemanfaatan, butir soal yang terpilih dapat dilihat saja dan dapat dikonvert ke word. Pengguna dibuatkan username dan password. Model ini disajikan pada Gambar 5.3.



Gambar 5.3. Sistem Bank Soal yang Dikembangkan

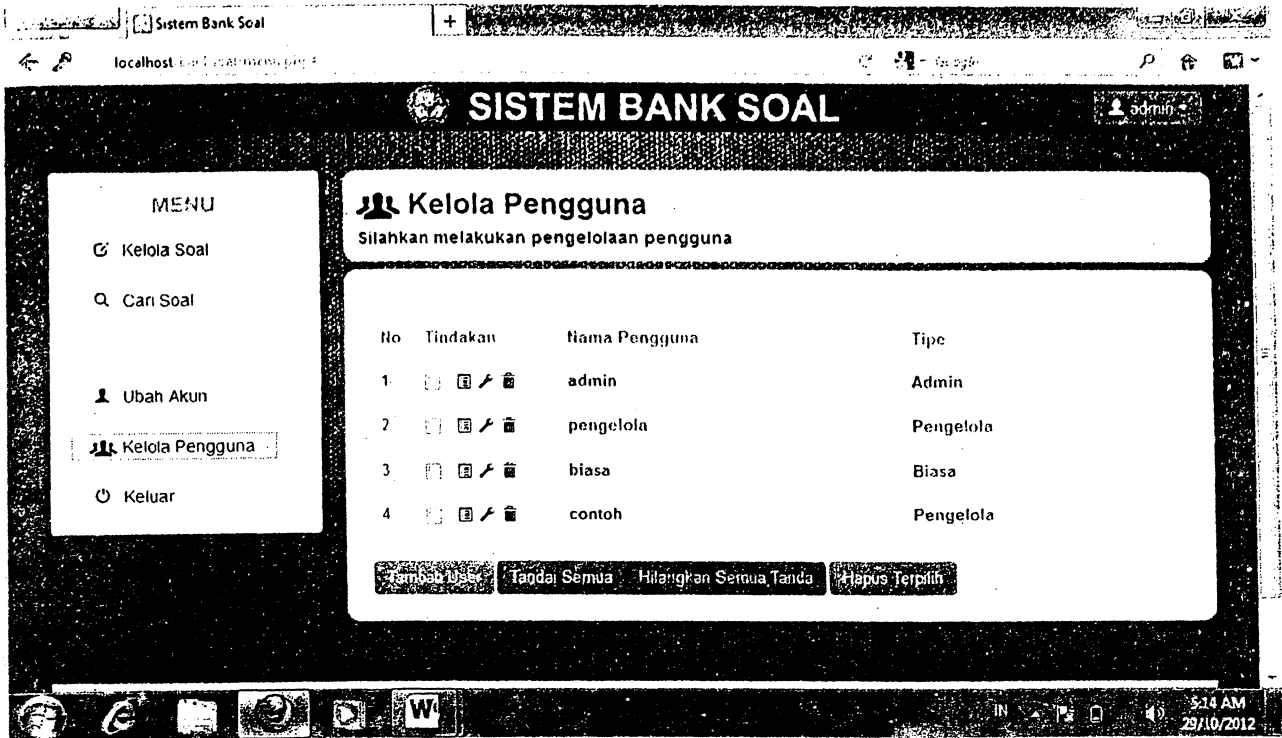
Sistem bank soal yang dikembangkan berbasis teknologi informasi, kemudian ditampilkan dalam web dengan basis program MySQL. Tampilan awal disajikan pada Gambar 5.4, log in dengan menggunakan username dan password disajikan pada Gambar 5.5. User dapat mengubah identitas yang disajikan pada Gambar 5.6. Menu mencari soal disajikan pada Gambar 5.7 dan menu mengelola butir disajikan pada Gambar 5.8. Menu yang dipilih kemudian dimasukkan ke keranjang (Gambar 5.9 dan 5.10) yang selanjutnya dapat dilihat saja, dicetak, atau dikonvert ke dokumen (*.doc) untuk diedit dan digunakan (Gambar 5.11).



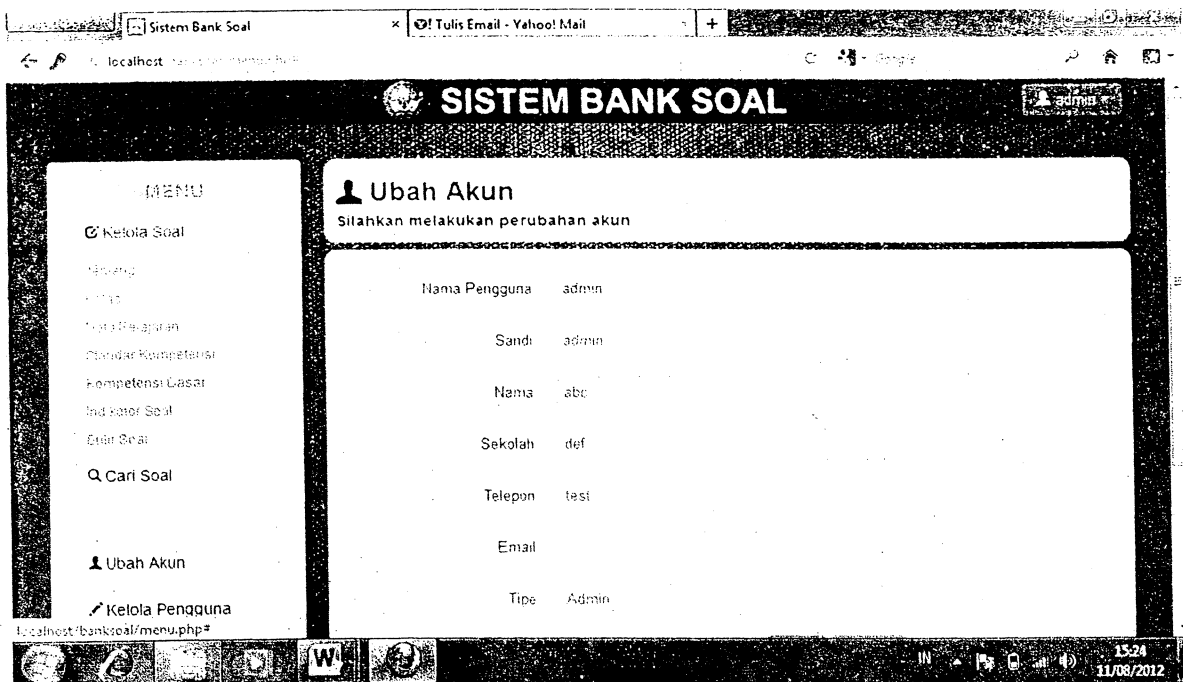
Gambar 5.4. Tampilan Awal Sistem Bank Soal



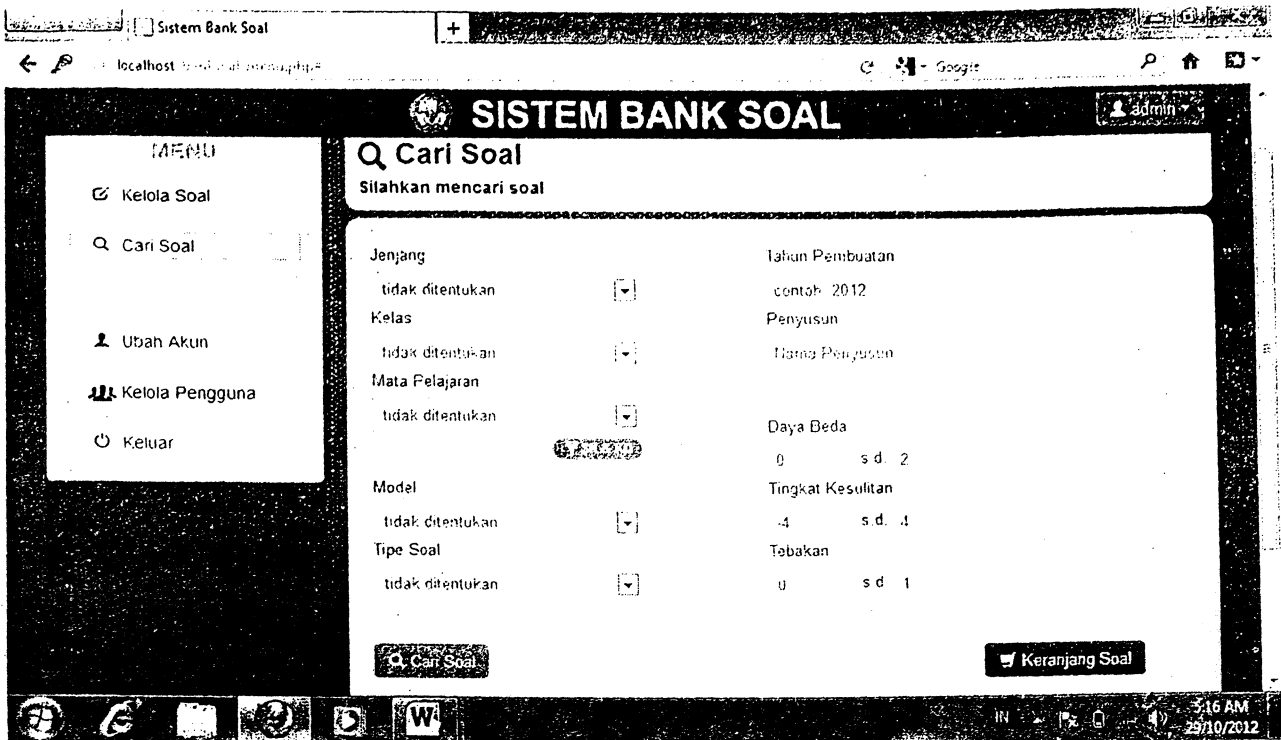
Gambar 5.5. Masuk ke Sistem Bank Soal Menggunakan Username dan Password



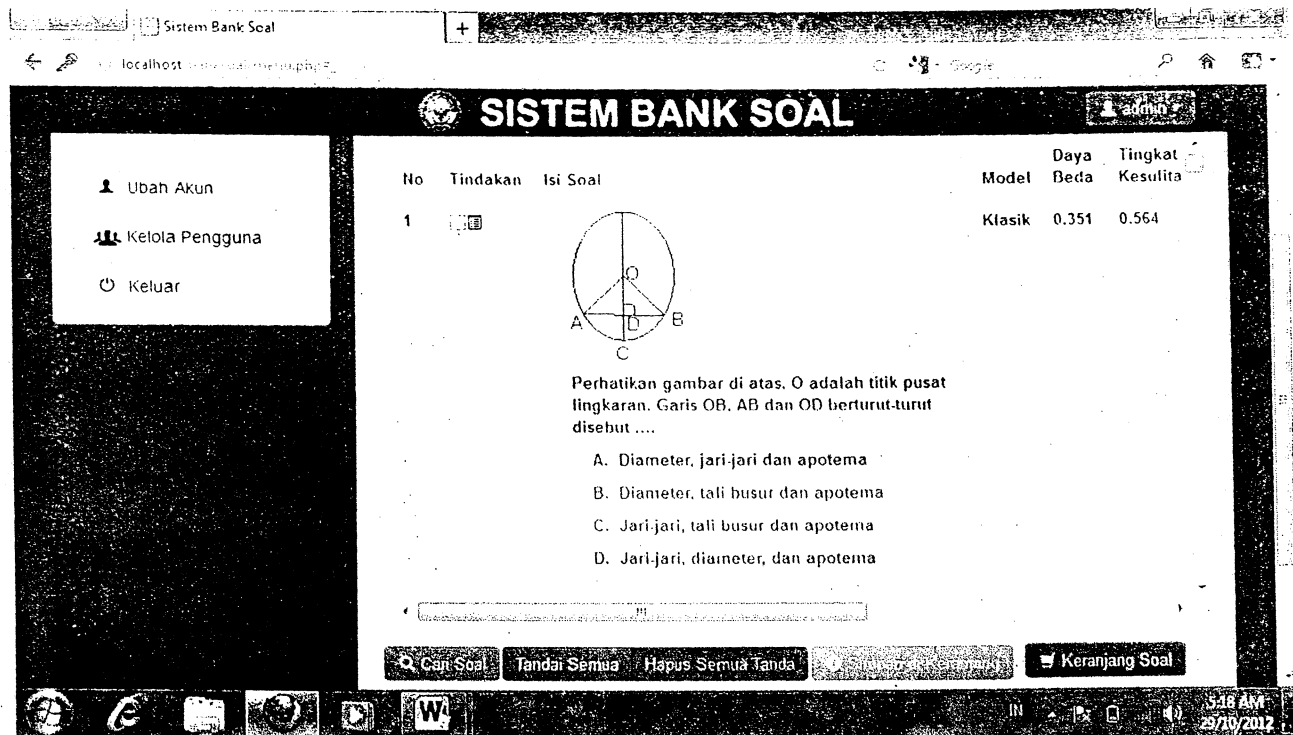
Gambar 5.6. Menu Mengelola Pengguna



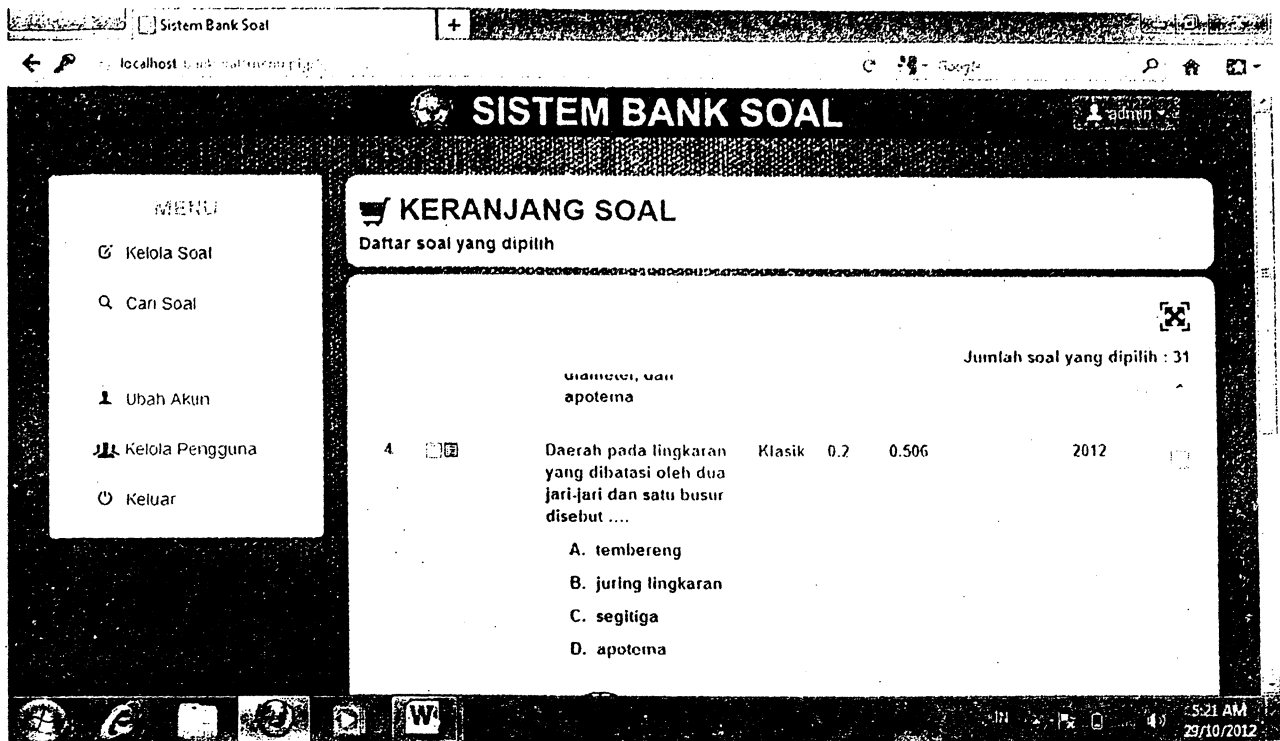
Gambar 5.7. Mengubah Identitas User



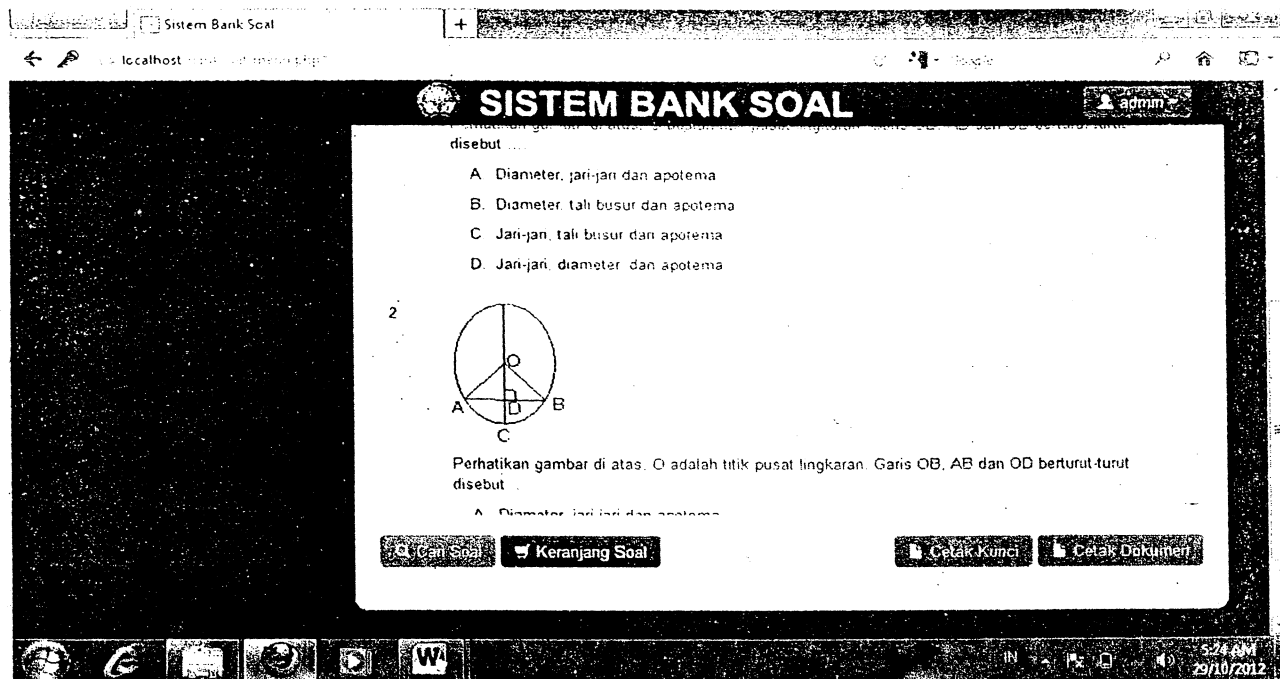
Gambar 5.8. Mencari dan Memilih Butir dalam Bank Soal



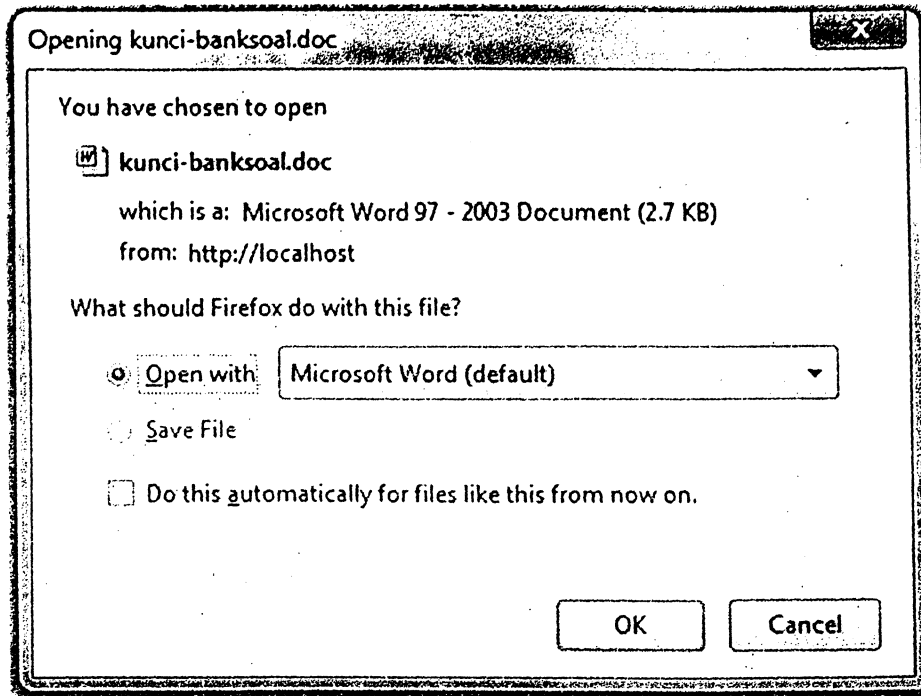
Gambar 5.9. Tampilan Soal dalam Sistem Bank Soal



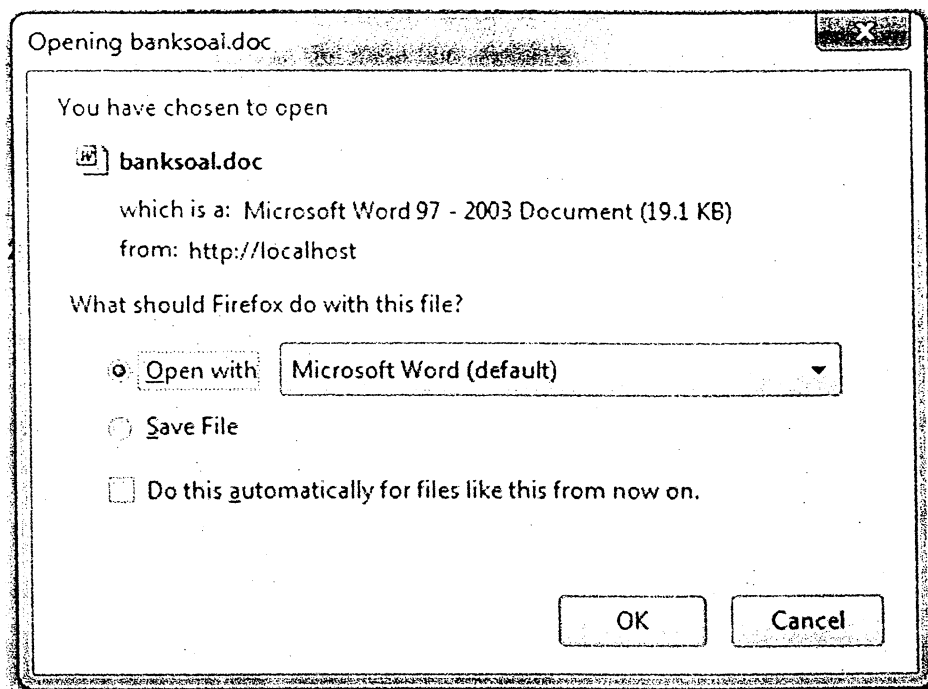
Gambar 5.10. Soal yang Terpilih di Dalam Keranjang Soal



Gambar 5.11. Tindak Lanjut Keranjang Soal



Gambar 5.12. Mencetak Kunci Jawaban ke Format *.doc



Gambar 5.13. Mencetak Soal ke Format *.doc

Selama dimanfaatkan, bank soal memerlukan penambahan butir terus menerus. Untuk penambahan ini, butir-butir yang ada perlu diketahui karakteristiknya dahulu (dikaliberasi) dan kemudian disetarakan dengan butir-butir yang telah ada melalui prosedur *equating*. Demikian pula butir-butir yang telah

dipakai, dapat pula dibuang (*delete*) atau disiman selama periode waktu yang cukup lama kemudian dipakai lagi.

* * *

Daftar Kepustakaan

Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ : Prentice Hall.

Cronbach, L.J. 1970. *Essential of psychological testing* (4th. ed.). New York: Harper & Row Publishers.

Gronlund, N.E. (1976). *Measurement and evaluation in teaching*. New York : Macmillan Publishing Co.

Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.

Heri Retnawati & Samsul Hadi. (2013). pengembangan sistem bank soal untuk ujian akhir daerah di era otonomi daerah dan desentralisasi. *Laporan Penelitian*. Universitas Negeri Yogyakarta.

Jahja Umar. (1999). Item banking. Dalam Masters, G.N. dan Keeves, J.P. (Ed). *Advances in Measurement in Educational Research and Assessment*. New York : Pergamon.

Thorndike, R.L. (1982). *Applied Psychometrics*. Boston : Houghton Mifflin.

BAB 6

MERAKIT PERANGKAT TES DENGAN MEMANFAATKAN NILAI FUGSI INFORMASI



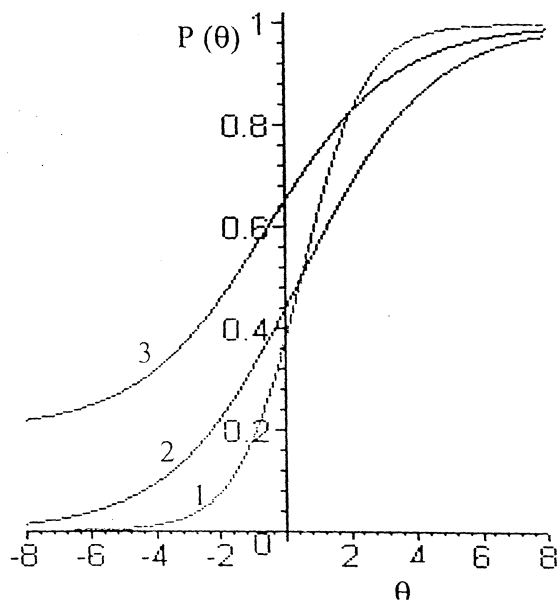
Salah satu penerapan dari teori respons butir adalah perakitan perangkat tes. Perangkat tes ini dapat dirakit menggunakan bank butir atau sering disebut juga bank soal, yang telah dikembangkan sesuai dengan tujuan tes. Misalnya untuk akhir kenaikan kelas mata pelajaran matematika SMP, diperlukan butir-butir dari bank soal matematika SMP.

Bank soal tidak hanya sekumpulan butir soal yang diketahui karakteristiknya. Namun bank soal merupakan sistem yang mengorganisir penyimpanan butir baik dari perangkat tes yang siap digunakan, pemanfaatan butir, dan penghapusan butir yang sudah tidak dapat digunakan lagi. Tentu saja pemanfaatan butir-butir ini disesuaikan dengan tujuan tes dan tujuan pengembangan bank soal.

Dengan teori respons butir unidimensi, bank butir dapat dikembangkan dengan menggunakan 3 model, yaitu model logistic 1 parameter (1PL), 2 parameter dan 3 parameter. Parameter butir tersebut yaitu tingkat kesulitan (1PL), tingkat kesulitan dan daya pembeda (2PL), tingkat kesulitan, daya pembeda, dan parameter tebakan semu (*pseudo guessing*). Dengan parameter-parameter ini, kemampuan peserta tes (θ) dapat diestimasi setelah mengerjakan serangkaian butir tes, dan hubungan probabilitas menjawab benar dengan parameter butir model 3PL disajikan secara matematis pada persamaan 1 (Hambleton, Swaminathan, dan Rogers, 1991: 17, Hambleton, dan Swaminathan, 1985 : 49, Van der Linden dan Hambleton, 1997: 13).

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \dots\dots\dots (7.1)$$

Persamaan tersebut merupakan persamaan dengan a daya pembeda, b tingkat kesulitan, dan c parameter tebakan semu. Model 1PL merupakan kasus khusus dari 3PL dengan a=1 dan c=0, sedang model 2PL merupakan kasus khusus dari model 3PL dengan c=0. Sebagai ilustrasi, gambar 6.1 merupakan kurva karakteristik butir 1 (a=1, b=0,5, c=0), butir 2(a=0,5, b=0,5, c=0) dan butir 3 (a=0,5, b=0,5, c=0,2).



Gambar 6.1. kurva karakteristik butir model 3P, dengan butir 1 (a=1, b=0,5, c=0), butir 2(a=0,5, b=0,5, c=0) dan butir 3 (a=0,5, b=0,5, c=0,2)

Nilai kemampuan peserta (θ) terletak di antara -4 dan $+4$, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Secara teoretis, nilai b_i terletak di antara $--$ dan $+-$. Suatu butir dikatakan baik jika nilai ini berkisar antara -2 dan $+2$ (Hambleton dan Swaminathan, 1985: 107). Jika nilai b_i mendekati -2 , maka indeks kesukaran butir sangat rendah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesukaran butir sangat tinggi untuk suatu kelompok peserta.

Parameter a_i merupakan daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik, a_i merupakan kemiringan (*slope*) dari kurva di titik b_i pada skala kemampuan tertentu. Karena merupakan kemiringan, diperoleh semakin besar kemiringannya, maka

semakin besar daya pembeda butir tersebut. Secara teoretis, nilai a_i ini terletak antara $-\infty$ dan $+\infty$. Pada pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dengan kemampuan yang diukur, dan a_i terletak antara 0 dan 2 (Hambleton dan Swaminathan, 1985: 37).

Peluang menjawab benar dengan memberikan jawaban tebakan. semu dilambangkan dengan c_i , yang disebut dengan *tebakan semu*. Parameter ini memberikan suatu kemungkinan asimtot bawah yang tidak nol (*nonzero lower asymptote*) pada kurva karakteristik butir (ICC). Parameter ini menggambarkan probabilitas peserta dengan kemampuan rendah menjawab dengan benar pada suatu butir yang mempunyai indeks kesukaran yang tidak sesuai dengan kemampuan peserta tersebut. Besarnya harga c_i diasumsikan lebih kecil daripada nilai yang akan dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir. Pada suatu butir tes, nilai c_i ini berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai c_i tidak lebih dari $1/k$, dengan k banyaknya pilihan (Hullin, 1983: 36).

Terkait dengan karakteristik butir dan bank soal, hanya butir yang karakteristiknya baik yang dapat disimpan dalam bank soal. Untuk bank soal yang dikembangkan dengan teori respons butir, tingkat kesulitan dapat merentang dari yang mudah ke yang sulit, daya pembeda antara 0-2, dan untuk butir pilihan ganda, c tidak melebihi $1/k$ dengan k banyaknya pilihan. Butir-butir dari bank soal ini dapat dimanfaatkan dengan dirakit menjadi perangkat tes yang bersesuaian dengan peruntukan bank butir.

Ada beberapa cara memilih butir dari bank soal untuk dirakit menjadi perangkat tes. Cara pertama merupakan cara yang paling sering digunakan, dengan memperhatikan isi soal. Dengan melihat indikator butir, dapat dirakit perangkat tes. Cara kedua yakni dengan melihat isi dan tingkat kesulitan. Cara ini mempertimbangkan informasi tingkat kesulitan yang tersedia dalam bank soal. Cara ketiga dengan memanfaatkan nilai fungsi informasi. Cara ini belum lazim digunakan, karena menggunakan pendekatan teori respons butir. Terkait dengan hal tersebut, pada bagian ini dibahas pemanfaatan nilai fungsi informasi untuk merakit tes disertai dengan simulasinya.

Fungsi informasi butir (*item information functions*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat soal dan menyatakan kekuatan atau sumbangan butir soal dalam mengungkap kemampuan laten (*latent trait*) yang diukur dengan tes tersebut. Dengan fungsi informasi butir diketahui butir mana yang

cocok dengan model sehingga membantu dalam seleksi butir soal. Secara matematis, fungsi informasi butir didefinisikan sebagai berikut.

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \dots\dots\dots (6.2)$$

keterangan :

i : 1,2,3,...,n

$I_i(\theta)$: fungsi informasi butir ke-i

$P_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

$P'_i(\theta)$: turunan fungsi $P_i(\theta)$ terhadap θ

$Q_i(\theta)$: peluang peserta dengan kemampuan θ menjawab salah butir i

Fungsi informasi butir untuk model logistik tiga parameter dinyatakan oleh Birnbaum (Hambleton & Swaminathan, 1985: 107) dalam persamaan berikut.

$$I_i(\theta) = \frac{2,89a_i^2(1-c_i)}{[(c_i + \exp(Da_i(\theta - b_i)))][1 + \exp(-Da_i(\theta - b_i))]^2} \dots\dots\dots(6.3)$$

keterangan :

$I_i(\theta)$: fungsi informasi butir i

θ : tingkat kemampuan subjek

a_i : parameter daya beda dari butir ke-i

b_i : parameter indeks kesukaran butir ke-i

c_i : indeks tebakan semu (*pseudoguessing*) butir ke-i

e : bilangan natural yang nilainya mendekati 2,718

Berdasarkan persamaan fungsi informasi di atas, maka fungsi informasi memenuhi sifat: (1) pada respons butir model logistik, fungsi informasi butir mendekati maksimal ketika nilai b_i mendekati θ . Pada model logistik tiga parameter nilai maksimal dicapai ketika θ terletak sedikit di atas b_i dan indeks tebakan semu butir menurun; (2) fungsi informasi secara keseluruhan meningkat jika parameter daya beda meningkat. De Gruijter & Van der Camp (2005: 118) menyatakan bahwa nilai fungsi informasi butir dan juga nilai fungsi informasi tes, bergantung pada kemampuan laten.

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir-butir tes tersebut (Hambleton & Swaminathan, 1985: 94). Berkaitan dengan hal ini, nilai fungsi

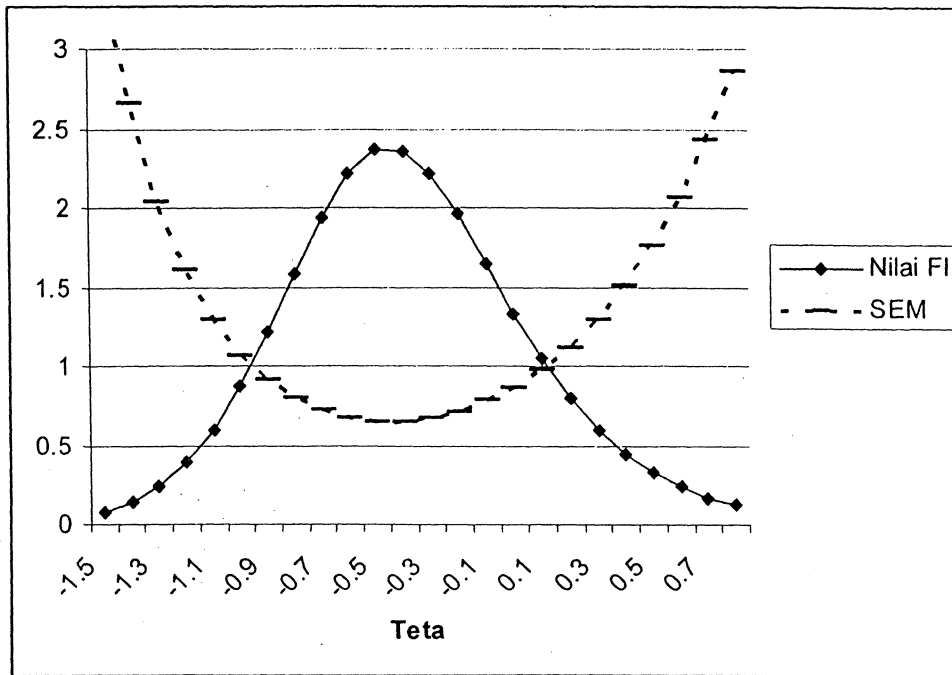
informasi perangkat tes akan tinggi jika butir-butir penyusun tes mempunyai fungsi informasi yang tinggi pula. Fungsi informasi perangkat tes ($I(\theta)$) secara matematis dapat didefinisikan sebagai berikut.

$$I(\theta) = \sum_{i=1}^n I_i(\theta) \dots\dots\dots (6.4)$$

Nilai-nilai indeks parameter butir dan kemampuan peserta merupakan hasil estimasi. Karena merupakan hasil estimasi, maka kebenarannya bersifat probabilistik dan tidak terlepas dengan kesalahan pengukuran. Dalam teori respons butir, kesalahan pengukuran standar (*Standard Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan SEM mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka SEM semakin kecil atau sebaliknya (Hambleton, Swaminathan, & Rogers, 1991, 94). Jika nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ dan nilai estimasi SEM dinyatakan dengan $SEM(\hat{\theta})$, maka hubungan keduanya, menurut Hambleton, Swaminathan, & Rogers (1991 : 94) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (6.5)$$

Sebagai ilustrasi, pada Gambar 6.2 disajikan grafik nilai fungsi informasi butir dan kesalahan pengukuran standar suatu butir dengan parameter $a=2$, $b=-0,5$, dan $c=0,1$.

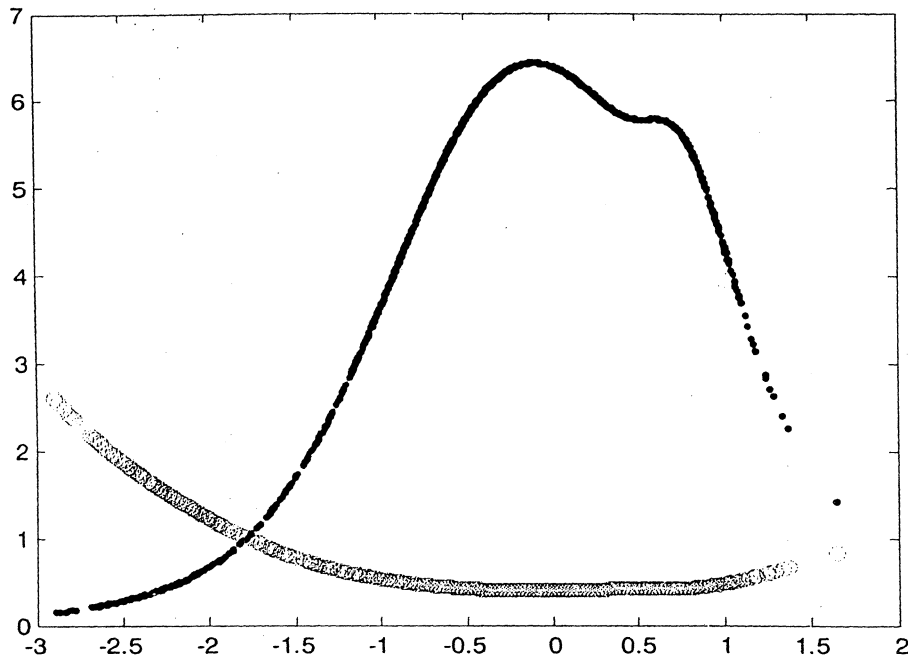


Gambar 6.2

Grafik Nilai Fungsi Informasi Butir dan Kesalahan Pengukuran Standar Butir dengan Parameter $a=2$, $b=-0,5$, dan $c=0,1$

Interpretasi dari Gambar 6.2 tersebut sebagai berikut. Nilai fungsi informasi dari kemampuan \sim naik, mencapai nilai maksimum, kemudian menurun sampai \sim . Sedangkan kesalahan pengukuran sebaliknya, menurun, mencapai nilai minimum, kemudian naik kembali. Kedua grafik fungsi ini bertemu pada skala kemampuan $-0,9$ dan $+0,2$. Di antara dua kemampuan ini, butir memiliki nilai fungsi informasi lebih tinggi dibanding kesalahan pengukurannya. Sebaliknya, ketika skala kemampuan kurang dari $-0,9$ dan lebih dari $+0,2$, butir memiliki kesalahan pengukuran dibandingkan dengan informasi yang diberikanya.

Demikian pula halnya dengan tes, dengan nilai fungsi informasi merupakan jumlahan dari nilai informasi butir penyusunnya. Sebagai ilustrasi pada Gambar 6.3, yang menyajikan nilai fungsi informasi dari suatu tes matematika SMP skala nasional dengan 40 butir. Ini berarti bahwa perangkat tes ini memiliki nilai fungsi informasi yang lebih tinggi dibandingkan dengan kesalahan pengukuran pada rentang $-1,7$ sampai $+1,8$. Dengan demikian, perangkat tes ini sesuai untuk peserta tes pada rentang kemampuan tersebut. Jika skala kemampuan kurang dari $-1,7$ dan lebih dari $+1,8$ kesalahan pengukuran lebih besar dibandingkan dengan nilai fungsi informasinya.



Gambar 6.3
 Nilai Fungsi Informasi Perangkat Tes Matematika

Seiring dengan perkembangan ilmu dan teknologi, perakitan perangkat menyesuaikan dengan kemampuan peserta tes, misalnya *computerized adaptive testing* (CAT). Hal ini dilakukan untuk memperoleh nilai fungsi informasi tes sebesar-besarnya dan meminimalkan kesalahan pengukuran. Dengan demikian, dapat dirakit perangkat tes yang sesuai dengan kemampuan, misalnya untuk peserta kelompok kemampuan tinggi, sedang, dan rendah. Hal ini mengantisipasi kasus peserta dengan kemampuan rendah dan mengerjakan perangkat untuk kemampuan sedang, dapat diperoleh kesalahan pengukuran yang lebih tinggi dibanding informasi, sehingga perlu dirakit tes khusus yang disesuaikan dengan kemampuan peserta.

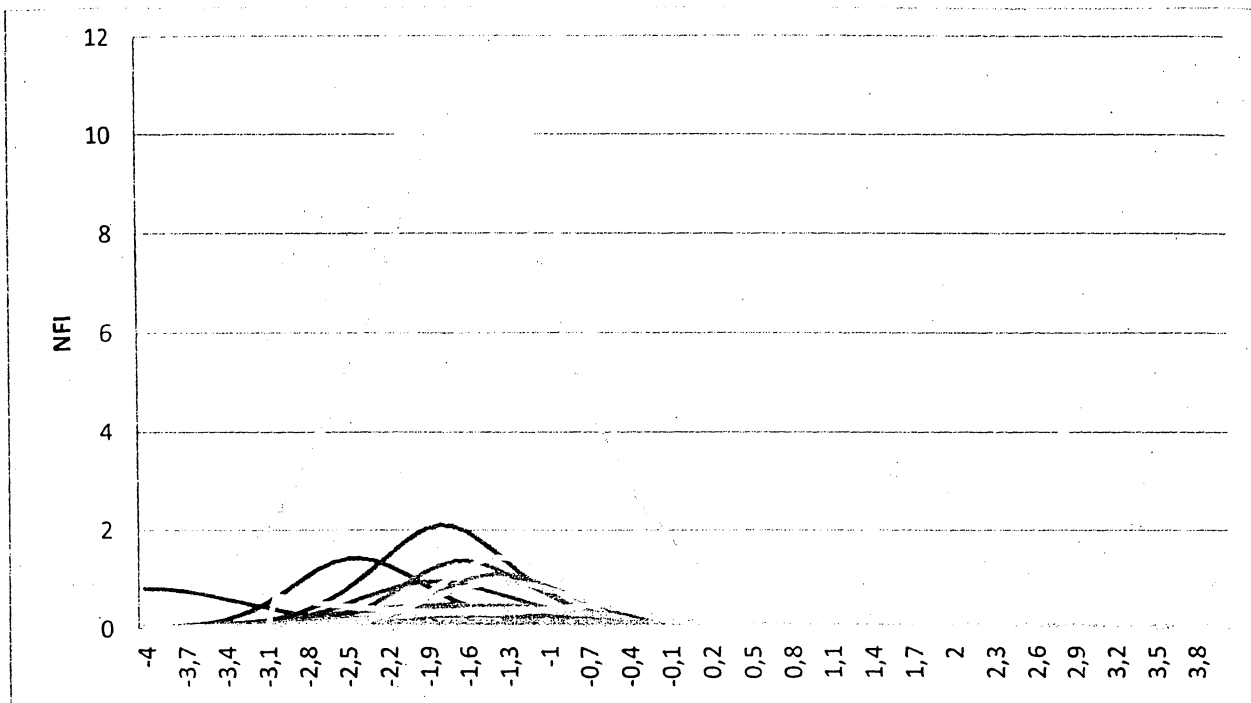
Pada perakitan tes ini, modal awal yang diperlukan adalah tersedianya bank butir. Bank butir merupakan suatu sistem pengelolaan butir yang menyimpan butir-butir yang baik dan menyediakan mekanisme penggunaannya. Agar pengguna lebih leluasa menggunakan butir, dalam bank butir sebaiknya bank ini menyimpan cukup banyak butir yang baik dengan tingkat kesulitan yang bervariasi.

Adapun langkah-langkah merakit perangkat dengan menggunakan bank butir yakni:

1. Menentukan SEM yang diinginkan
Misalnya pakar yang akan menggunakan tes menginginkan kesalahan pengukuran sebesar 0,316228.
2. Menentukan NFI target dan kemampuan calon peserta tes
Dengan menggunakan persamaan 5, dapat dihitung nilai fungsi informasi target sebesar 10, dan misalnya akan digunakan untuk calon peserta dengan kemampuan -2,5 sampai -0,8.
3. Menggambar NFI target
Selanjutnya nilai fungsi target tersebut digambarkan dalam koordinat kartesius, dengan sumbu mendatar kemampuan dan sumbu tegak nilai fungsi informasi.
4. Memasukkan Butir-butir dari bank yang menentukan tercapainya nilai fungsi informasi target
Butir-butir yang mendukung dipilih, dengan jumlah nilai fungsi informasi mendekati nilai fungsi informasi target.
5. Butir-butir pada langkah nomor 4 tersebut yang kemudian disusun menjadi perangkat tes, dengan menggunakan criteria penyusunan tes yang baik, misalnya diurutkan dari yang paling mudah ke yang paling sulit.

Sebagai bahan simulasi, dibangkitkan 200 butir soal dengan model 3PL, dengan tingkat kesulitan berdistribusi seragam pada rentang (-3,+3), daya pembeda pada rentang (0,2) dan tebakan semu kurang dari 0,25 dengan WINGEN. Selanjutnya dengan langkah-langkah merakit perangkat tes dari bank butir tersebut, disimulasikan perakitan 3 buah tes untuk kelompok kemampuan level bawah, sedang, dan kemampuan atas.

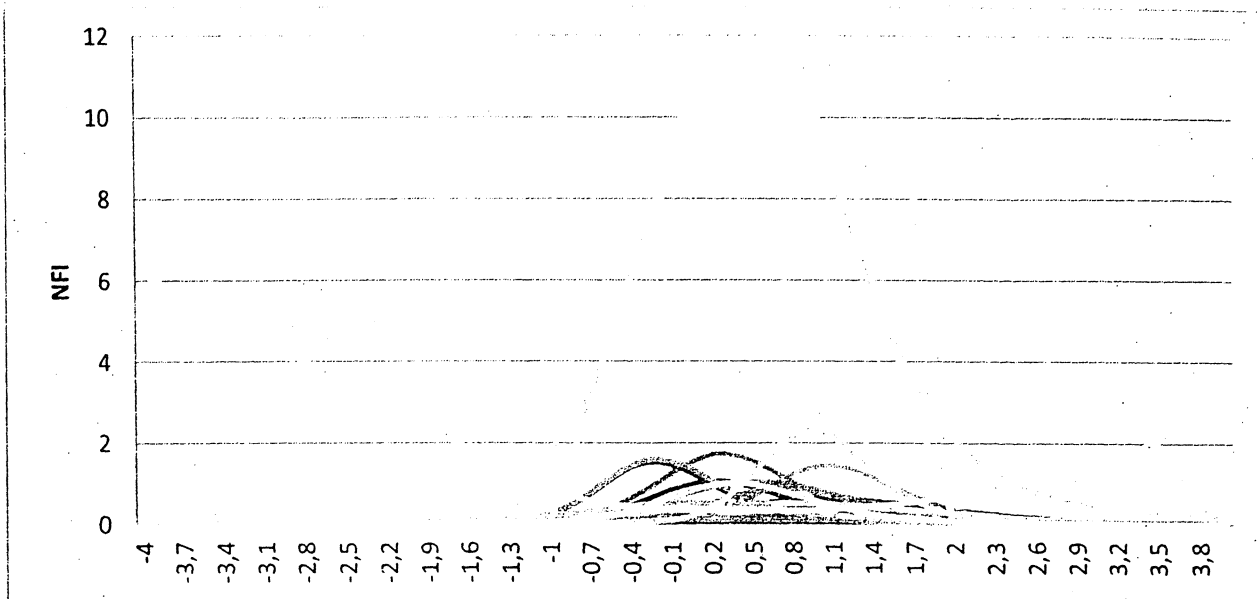
Pada kelompok bawah, dengan 20 butir dan rerata tingkat kesulitan -1,84 diperoleh nilai fungsi informasi yang mendekati target nilai fungsi informasi. Perangkat tes ini cocok untuk peserta dengan skala kemampuan (-4, -0,3). Hasil selanjutnya digambarkan pada Gambar 6.4.



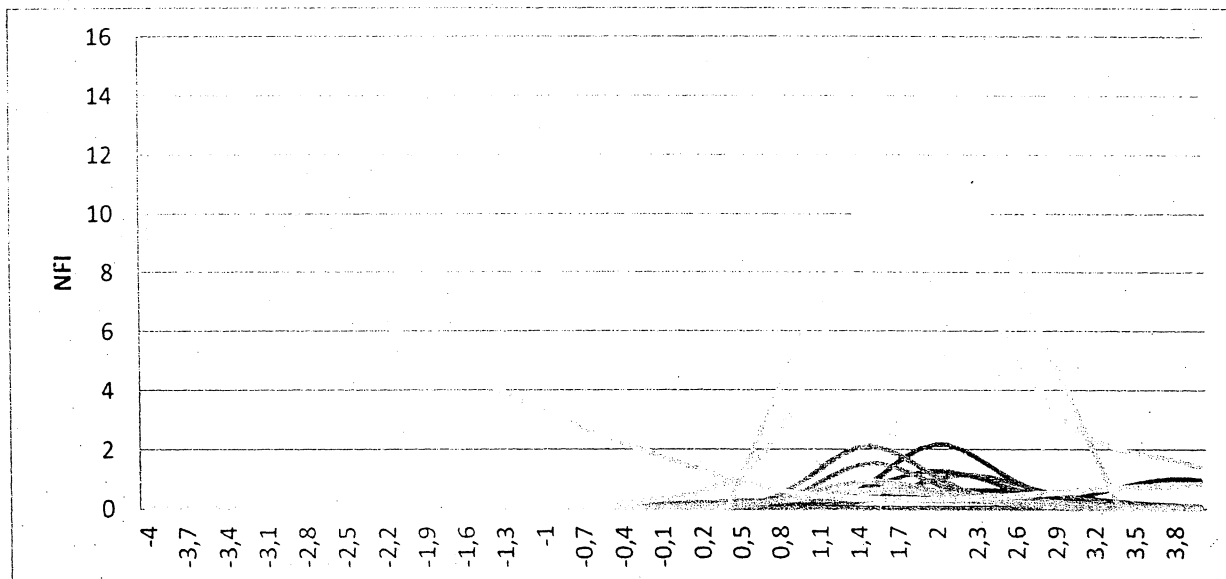
Gambar 6.4
 Nilai fungsi informasi target dan nilai fungsi informasi tes
 dan butir yang menyusunnya
 Untuk peserta dengan kemampuan rendah

Untuk kelompok sedang, dengan 18 butir dan rerata tingkat kesulitan +0,315 diperoleh nilai fungsi informasi yang mendekati target nilai fungsi informasi. Perangkat tes ini cocok untuk peserta dengan skala kemampuan (-1,2, +2,4). Hasil selanjutnya digambarkan pada Gambar 6.5.

Adapun pada kelompok atas, dengan 1 butir dan rerata tingkat kesulitan +1,84 diperoleh nilai fungsi informasi yang mendekati target nilai fungsi informasi. Perangkat tes ini cocok untuk peserta dengan skala kemampuan (-0,4, +4,0). Hasil selanjutnya digambarkan pada Gambar 6.6.



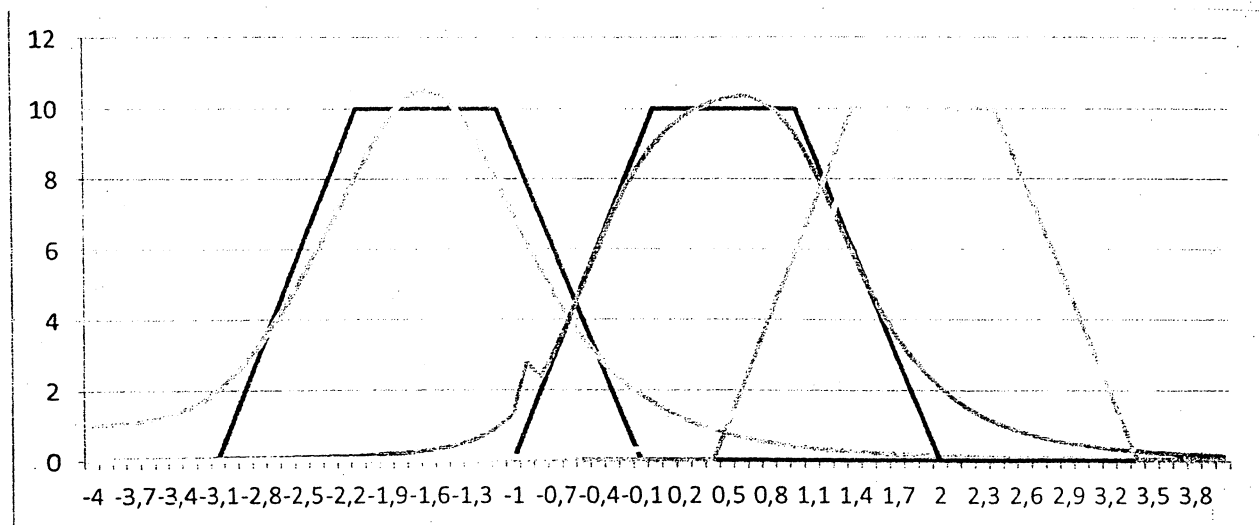
Gambar 6.5
 Nilai fungsi informasi target dan nilai fungsi informasi tes
 dan butir yang menyusunnya
 Untuk peserta dengan kemampuan sedang



Gambar 6.6
 Nilai fungsi informasi target dan nilai fungsi informasi tes dan butir yang menyusunnya
 Untuk peserta dengan kemampuan tinggi

Jika nilai fungsi informasi baik target maupun hasil disatukan, diperoleh bahwa nilai fungsi informasi ketiga tes ini merentang dari skala kemampuan -4,0 samapai dengan +4,0. Hal ini menunjukkan bahwa ketiga perangkat tes dapat digunakan untuk

mengukur kemampuan seluruh siswa dengan mendapatkan nilai fungsi informasi yang baik sehingga dapat meminimalkan kesalahan pengukuran. Hasil selengkapnya disajikan pada Gambar 6.7.



Gambar 6.7
 Nilai fungsi informasi target dan nilai fungsi informasi tes
 Untuk peserta dengan kemampuan rendah, sedang, dan tinggi

Dengan menggunakan bank soal yang memuat 200 butir soal yang parameternya disimulasikan, dapat disusun perangkat tes untuk kemampuan peserta yang berbeda-beda dengan nilai fungsi informasi maksimum yang hampir sama. Misalnya pada kasus ini untuk peserta pada level kemampuan rendah, sedang, dan tinggi. Perangkat ini kemudian dapat digunakan sesuai dengan tujuan pengembangan tes.

Meskipun merakit tes ini didasarkan pada nilai fungsi informasi butir yang juga melibatkan tingkat kesulitan butir pada teori respons butir unidimensi, ada beberapa kelemahan yang terdapat pada prosesnya. Dalam merakit tes, tetap perlu diperhatikan isi/substansi butir, yang dapat diketahui dari indicator soal. Jadi dalam merakit tes, tidak semata-mata mengejar ketercapaian nilai informasi butir saja, namun tetap mempertahankan validitas isinya.

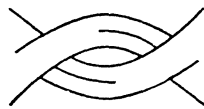
Dengan melihat kasus perakitan tes pada level rendah, sedang, dan tinggi, diperoleh bahwa untuk mengejar nilai fungsi informasi target, diperoleh jumlah butir yang berbeda. Dengan adanya perbedaan banyaknya butir tiap level, berimplikasi pada waktu tes. Perangkat tes dengan butir lebih banyak tentunya memerlukan waktu yang lebih lama. Hal ini perlu diperhitungkan pada administrasi tesnya.

Daftar Pustaka

- Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.
- Hullin, C. L., et al. (1983). *Item response theory : Application to psychological measurement*. Homewood, IL : Dow Jones-Irwin.
- Van der Linden, W.J. dan Hambleton, R.K. (1997). *Item response theory:brief history, common models and extentions*. Dalam Van der Linden, W.J. dan Hambleton, R.K. (Eds). *Handbook of item response theory*. New York : Springer.

BAB 7

PENYETARAAN (EQUATING) DAN CONCORDANCE



A. Pengertian dan Asumsi Menghubungkan Tes-Tes

Skor-skor pada asesmen hasil pengukuran pendidikan dapat disetarakan secara statistik, dari satu unit asesmen ke unit asesmen yang lain, atau keduanya dapat dinyatakan dalam sebuah skala skor yang biasa. Cara ini disebut dengan menghubungkan dua tes (*linking*). Istilah *linking* merujuk pada sebuah hubungan antar skor dari dua tes. Seringkali dua tes yang dikaitkan ini mengukur konstruk yang sama, namun untuk kepentingan tertentu, mengaitkan dua tes yang berbeda konstruksinya.

Istilah *concordance* merujuk pada mengaitkan skor-skor pada asesmen yang mengukur konstruk yang sama dan skor-skor ini dihubungkan untuk suatu kepentingan. Sebagai contoh, calon mahasiswa dapat menempuh tes SAT I atau tes ACT. Sebuah perguruan tinggi mewajibkan siswa memiliki skor komposit 1200 untuk Verbal dan Matematika. Jika perguruan tinggi menerima siswa yang menempuh ACT, perguruan tinggi perlu mengetahui skor komposit ACT yang sebanding dengan skor minimum SAT I (Brennan & Kolen, 2004).

Flanagan menggunakan istilah *comparability* untuk menunjukkan tes-tes yang terskalakan yang memiliki distribusi skor yang sama dalam populasi nyata peserta-peserta tes. Lebih jauh, tes-tes yang skor-skornya dapat dibandingkan jika dalam pemakaian, tes-

tes tersebut dapat secara komplet saling dipertukarkan (*interchangeable*) dalam penggunaannya (Kolen, 2004). Keadaan saling dipertukarkan ini terjadi jika tes-tes tersebut memiliki akurasi yang sama dan didesain sebagai bentuk-bentuk yang ekuivalen (*equivalent forms*). Flanagan menekankan pada skor bentuk multipel, tes yang dapat saling dipertukarkan jika terkonstruksi sama dan skala skor harus digunakan untuk menyatakan skor-skoranya. Mengenai hubungan *comparability* dengan reliabilitas, Flanagan menyatakan bahwa untuk menetapkan skor-skor yang dapat dibandingkan, distribusi skor sebenarnya (*true score*) seharusnya sama pada pengukuran-pengukuran dari tes-tes tersebut. Jika reliabilitas pengukuran sama untuk kedua tes, hasil yang sama akan diperoleh jika distribusi dari nilai yang ada diperbandingkan. Pada tes yang berbeda, tidak perlu memberikan skor-skor yang terbandingkan. Hal ini disebabkan karena regresi pada kedua tes tidak simetri, dan ini mengakibatkan kekurangcukupan pada interpretasi.

Untuk menghubungkan skor-skor tes yang memiliki bentuk-bentuk lain yang dibangun dengan spesifikasi yang sama, Angoff menggunakan istilah penyetaraan (*equating*) (Brennan & Kolen, 2004). Sedangkan istilah yang digunakan untuk menyatakan hubungan antara dua tes yang memiliki konstruk yang sama tetapi berbeda dalam tingkat kesulitan dan reliabilitas, digunakan istilah kalibrasi (*calibration*) dan *comparability* digunakan untuk menyatakan hubungan tes-tes yang secara konstruk berbeda. *Comparability* merupakan keadaan ketika skor dari tes-tes yang berbeda diskalakan ke sebuah skor tes memiliki sebuah distribusi biasa untuk menilai kekuatan dan kelemahan peserta tes terhadap suatu grup yang dijadikan acuan. Ketika peserta tes memilih untuk menempuh suatu tes, tingkat kemampuannya akan berbeda jika menempuh tes yang lain. Skor-skor tes yang dibandingkan menjadi tidak unik lagi. Ketidak unikan ini disebabkan oleh karena pada kenyataannya alat-alat ukur memiliki fungsi yang berbeda, sehingga tidak ada konversi tabel tunggal yang dapat diterapkan untuk semua grup. Menurut Angoff, kegunaan tabel skor yang dapat dibandingkan tergantung jawaban dari dua pertanyaan bagaimanakah kesamaan dari tes-tes untuk yang skornya dapat diperbandingkan dikembangkan dan bagaimanakah kesesuaian dari grup pada tabel skor yang dapat diperbandingkan didasarkan pada satu pertimbangan person atau grup tujuan tabel tersebut digunakan.

Mislevy dan Linn (Brennan & Kolen, 2004) mengembangkan kerangka kerja untuk menghubungkan skor tes-tes meliputi empat tipe hubungan statistik, yakni penyetaraan (*equating*), kalibrasi (*calibration*), moderasi statistik (*statistical moderation*) dan prediksi (*projection/prediction*). Seperti halnya Angoff, Mislevy/Linn menggunakan istilah *equating* untuk menghubungkan skor-skor yang bentuk-bentuknya berbeda pada asesmen-asesmen. Kalibrasi digunakan untuk menghubungkan skor-skor tes yang mengukur konstruk yang sama tetapi berbeda dalam reliabilitas atau tingkat kesulitannya. Proyeksi dan moderasi statistik digunakan untuk menghubungkan tes-tes yang mengukur konstruk yang berbeda, menggunakan metode regresi. Pada istilah moderasi statistika, yang istilah ini juga digunakan oleh Kevees, skor setiap tes dihubungkan dengan variabel ketiga yang disebut sebagai moderator. Grup-grup yang tesnya akan dihubungkan menempuh tes lain, yang merupakan variabel moderator. Mislevy dan Linn tidak membahas lebih lanjut hal-hal yang membedakan situasi yang mana yang mengukur konstruk yang sama dari situasi yang mengukur konstruk yang berbeda.

Freuer menggunakan istilah dan definisi yang sama dari istilahnya dan konsisten, seperti halnya yang dinyatakan oleh Mislevy dan Linn, yakni penyetaraan, kalibrasi, moderasi dan proyeksi. Freuer menambahkan, ada 5 faktor-faktor untuk dipertimbangkan tentang skor-skor yang akan dihubungkan, yakni :

- a. kesamaan isi, tingkat kesulitan dan format butir.
- b. dapat diperbandingkannya kesalahan pengukuran yang terkait dengan skor-skor,
- c. kondisi administrasi tes,
- d. kegunaan dibuatnya tes dan konsekuensinya,
- e. akurasi dan stabilitas dari penyetaraan, termasuk stabilitas atas subgrup dan peserta ujian-ujianya.

Dorans (2004) membedakan antara penghubungan skor-skor yang mengukur konstruk yang berbeda, dari skor-skor yang mengukur konstruk-konstruk yang sama. Penghubungan skor-skor pada tes-tes yang mengukur konstruk yang sama disebut dengan concordance. Konstruk yang sama ini diindikasikan dengan kesamaan isi, skor-skornya berkorelasi tinggi, dan hubungan antar peserta tes berbeda sedikit. Jika ada dua

tes-tes yang akan dihubungkan tidak memenuhi syarat concordance, maka untuk menghubungkannya dapat digunakan metode regresi.

Kolen dan Brennan mengajukan 4 situasi penghubungan skor-skor pada tes-tes, yaitu:

- a. inferensi (pada rentang apa kedua tes menggambarkan inferensi yang sama?)
- b. konstruk (pada rentang apa kedua tes mengukur konstruk yang sama?)
- c. populasi (pada rentang apa kedua tes didesain untuk digunakan pada populasi yang sama?)
- d. kondisi pengukuran (pada rentang apa kedua tes berada pada kondisi yang sama, misalnya panjang tes, formatnya, administrasinya, dan lain-lain?)

Sebagai contoh misalnya concordance pada ACT dan SAT I matematika didesain untuk digunakan pada populasi yang sama, diadministrasikan pada keadaan—kondisi yang sama, untuk inferensi yang sama, dengan konstruk yang sama pula.

Kesamaan konstruk berperan penting, dalam menentukan derajat menghubungkan skor-skor tes. Ada 3 derajat hubungan skor-skor tes-tes. Jika tes-tes tersebut secara statistik dan konseptual dapat saling menggantikan, maka hubungan dapat diketahui dengan penyetaraan (*equating*), jika sama distribusinya (mengukur konstruk yang sama) dengan *concordance*, dan jika kondisi untuk penyetaraan dan *concordance* tidak terpenuhi, digunakan prediksi skor harapan.

Tujuan dari penyetaraan adalah menghasilkan skor yang dapat saling menggantikan. Suatu ukuran dapat saling menggantikan dengan suatu ukuran yang lain jika ukuran tersebut diperoleh dari konstruk yang sama (misalkan panjangnya), dan sama ukurannya. *Concordance* akan terjadi jika mengukur konstruk yang sama dan terhubung dalam cara yang sama melintasi subpopulasi yang berbeda. Prediksi skor harapan merupakan penyetaraan ataupun *concordance* hanya ketika dua set skor-skor terkait dengan sempurna, yang hanya terjadi jika kedua set skor tersebut mengukur hal yang sama tanpa kesalahan dengan reliabilitas yang sama pula. Tidak seperti penyetaraan dan *concordance*, hubungan tidak bersifat simetris (fungsi konversi pada tes A ke tes B bukanlah fungsi inverse dari tes B ke tes A).

Lalu akan timbul pertanyaan, bagaimanakah menentukan hubungan dua tes, apakah penyetaraan, concordance ataukah prediksi skor harapan? Ada 3 faktor yang dapat menjawab permasalahan ini. Pertama, mengevaluasi kesamaan proses yang memproduksi skor-skor untuk melihat apakah konstruk yang diukur sama. Ini dapat ditempuh dengan mengevaluasi isi secara cermat. Kedua, mengakses kekuatan hubungan empiris antar skor yang dihubungkan. Prosedur ini dapat dilakukan dengan analisis faktor atau varians lain dari persamaan model struktural. Ketiga, mengakses derajat hubungan, yakni dengan mengetahui invarian lintas subpopulasi.

Dapat tidaknya disetarakan merupakan aspek penting dalam penyetaraan. Invarians populasi digunakan untuk membedakan derajat hubungan dua tes atau lebih, apakah merupakan penyetaraan (*equating*) atau concordance (Dorans, 2004). Jika ada invarian dalam populasi, digunakan penyetaraan, dan jika tidak ada invarian, digunakan concordance. Dorans memberikan formula *root means square difference* (RMSD) dan *root expected means square difference* (REMSD) untuk menentukan invarian populasi. Namun, jika untuk kasus khusus, misalnya variannya hanya 2, misalnya jenis keiamin (laki-laki dan perempuan), Dorans memberikan cara yang sederhana. Pada kasus ini REMSD sama dengan perbedaan dalam rata-rata skor standar antar grup dikalikan akar kuadrat dari ukuran relatif dari kedua grup. Dengan kata lain, jika hubungan tes-tes secara pasti paralel linear, dan pada kedua tes, perbedaan rata-rata antar grup laki-laki dan perempuan sama, maka populasi dikatakan invarian.

Penyetaraan didefinisikan khusus oleh Kolen & Brennan (1995). Penyetaraan merupakan proses statistic yang digunakan untuk mengatur skor pada dua perangkat tes atau lebih sehingga skor-skor pada perangkat tes dapat saling tukar. Tujuan penyetaraan adalah menempatkan skor dari dua tes pada skala yang sama. Menurut Hambleton, Swaminathan & Rogers (1991), skor dari dua perangkat tes dapat disetarakan dengan memperhatikan kondisi tes yang mengukur kemampuan yang berbeda tidak dapat disetarakan, skor dari tes-tes yang reliabilitasnya tidak sama tidak dapat disetarakan, dan skor pada tes-tes yang bervariasi tingkat kesukarannya dapat disetarakan. Adapun Lord (dalam Hambleton dan Swaminathan, 1985) menjelaskan ada 4 prinsip dasar penyetaraan. Prinsip pertama tersebut yakni prinsip kesetaraan (*equity*). Pada kondisi ini, setiap

kelompok peserta tes dengan kemampuan yang sama, kondisi distribusi frekuensi skor pada tes Y setelah ditransforasi sama dengan distribusi frekuensi skor pada tes X. Kedua, prinsip invariansi populasi, yang berarti hubungan penyetaraan pada transformasi tidak lagi memperhatikan kelompok populasi yang digunakan. Ketiga prinsip simetri, penyetaraan dapat dilakukan bolak-balik, tanpa memperhatikan tes mana yang diberi label X dan yang lainnya yang diberi label Y. Prinsip keempat adalah unidimensi, perangkat tes yang disetarakan mengukur kemampuan yang sama.

B. Jenis Penyetaraan dan *Concordance*

Penyetaraan dan *concordance* ada dua jenis, jenis horizontal dan dan vertikal. Pada jenis horizontal, dua skor tes atau lebih yang disetarakan merupakan tes-tes yang mengukur tingkat/kelas yang sama. Pada jenis ini, perangkat tes-perangkat tes yang diperbandingkan diberikan pada kelompok peserta tes yang memiliki distribusi kemampuan yang sama (Hambleton & Swaminathan, 1985).

Pada jenis vertikal, dua skor tes atau lebih yang disetarakan merupakan tes-tes yang mengukur tingkat/kelas yang berbeda, ada yang lebih tinggi atau lebih rendah dibandingkan lainnya. Penyetaraan vertikal (*vertical equating*) merupakan penyetaraan yang dilakukan terhadap dua instrumen tes atau lebih yang tingkat kesulitan butirnya berbeda, namun mengukur *trait* yang sama, dan distribusi skor peserta tes tidak komparabel sehingga skor-skor dari instrumen-instrumen tes tersebut dapat digunakan saling bertukar. Menurut Kolen & Brennan (1995:3), penyetaraan skor tes dengan *content* tidak berbeda dan kelompok peserta tes berasal dari tingkatan kelas berbeda, dan agar skor tes yang demikian dapat digunakan saling bertukar adalah penyetaraan vertikal. Menurut Hambleton & Swaminathan (1985:197) penyetaraan yang dilakukan terhadap beberapa instrumen tes dengan tingkat kesulitan soal berbeda dan distribusi kemampuan peserta tes juga berbeda disebut penyetaraan vertikal. Menurut Crocker & Algina (1986:473), penyetaraan vertikal dapat melibatkan dua atau lebih instrumen tes yang mengukur *trait* sama, namun tingkat kesulitannya berbeda.

Penyetaraan vertical (*vertical equating*), tes-tes yang disetarakan berbeda tingkat kesulitannya dan distribusi skor peserta tes tidak komparabel, serta bertujuan untuk membuat perbandingan kemampuan antarpeserta tes pada tingkatan yang berbeda. Hal ini menunjukkan bahwa penyetaraan vertikal memerlukan kelompok peserta tes berasal dari level kelas berbeda, tingkat kesulitannya berbeda, distribusi skor peserta tes berbeda, distribusi kemampuan peserta berbeda, namun mengukur *trait* yang sama (bersifat unidimensional). Beberapa asumsi dalam penyetaraan vertikal adalah (a) tes mengukur isi materi yang sama, (b) dimensi dasar yang diestimasi sama, (c) tes mengukur kemampuan (*latent trait*) yang unidimensi, dan (d) butir soal berbeda tingkat kesulitannya, tetapi bukan indeks diskriminasinya (Crocker & Algina, 1986:476). Pelanggaran terhadap asumsi dapat menimbulkan efek *bias* estimasi parameter θ . Asumsi-asumsi tersebut mendasari kegiatan penyetaraan vertikal, baik secara teoretis maupun praktis.

C. Desain Penyetaraan

Terdapat tiga desain dasar yang dapat digunakan untuk pengumpulan data dalam melakukan menghubungkan skor tes. Desain yang dimaksud adalah desain grup tunggal, desain grup ekuivalen, dan desain tes-anchor (Hambleton & Swaminathan, 1985: 198). Pemilihan desain yang digunakan dalam menghubungkan skor tes sangat tergantung pada situasi pelaksanaannya.

1. Desain Grup Tunggal

Pada desain grup tunggal, dua tes yang akan disetarakan diberikan kepada grup yang sama. Desain ini sederhana, tetapi tidak praktis dalam implementasinya, karena waktu pelaksanaan tes menjadi lama. Jika dua tes dilaksanakan berurutan tanpa diselingi waktu istirahat yang cukup, maka faktor latihan dan kelelahan berpengaruh pada estimasi parameter yang akhirnya berpengaruh pada hasil penyetaraan.

Untuk menghindari pengaruh urutan, latihan, dan kelelahan dalam pelaksanaan tes, dua tes diberikan dalam setiap dua grup yang ekuivalen secara random. Jika pelaksanaan tes ditempuh dengan cara demikian, maka desain ini

disebut desain grup tunggal dengan keseimbangan. Selanjutnya untuk menghindari kepenatan dalam pelaksanaan tes, dapat dikurangi dengan cara pelaksanaan tes dilakukan pada waktu dan lokasi yang berbeda. Desain grup tunggal kurang umum dalam praktik, karena pelaksanaan kedua tes tidak praktis, sehingga desain ini tidak disarankan dalam penerapannya.

2. Desain Grup Ekuivalen

Pada desain grup ekuivalen, dua tes yang akan disetarakan diberikan kepada dua grup yang ekuivalen (tidak identik) yang dipilih secara random dari populasi sama, kedua grup dianggap mempunyai tingkat kemampuan sama. Jadi desain ini hampir sama dengan desain grup tunggal. Dalam penerapannya, disamping mempunyai kelebihan juga mempunyai kelemahan.

Kelebihan desain ini adalah lebih praktis dan menghilangkan pengaruh latihan dan kepenatan. Kelemahan desain ini adalah adanya bias yang dihasilkan dari proses penyetaraan, karena grup-grup tersebut distribusi kemampuannya belum tentu sama. Untuk mengurangi bias yang ditimbulkan yang berkaitan dengan sampel, secara umum ukuran sampel yang besar diperlukan pada desain ini.

3. Desain Tes dengan Butir Bersama

Pada desain ini, dua tes yang akan disetarakan diberikan kepada dua grup yang berbeda. Masing-masing tes berisi satu set butir umum (*common item*) yang merupakan bagian dari butir-butir tes. Satu set butir umum yang merupakan bagian dari butir-butir tes disebut butir bersama (*common item* atau *anchor item*). Satu set butir bersama yang berdiri sendiri disebut butir bersama eksternal, sedangkan butir-butir bersama yang berada dalam kedua tes yang berbeda disebut butir bersama internal. Keuntungan menggunakan butir bersama internal dalam penyetaraan tes adalah masing-masing grup hanya menerima satu paket tes, waktu pelaksanaan tes lebih singkat, dan skor pada butir bersama ikut diperhitungkan dalam perhitungan skor total tes. Desain inilah yang sering digunakan untuk mengumpulkan data dalam penyetaraan. Penyetaraan tes yang

menggunakan butir bersama eksternal, pelaksanaannya dilakukan secara terpisah dengan tes total, skornya tidak diperhitungkan dalam perhitungan skor total.

Dalam penerapannya, desain penyetaraan dengan butir bersama melibatkan minimal dua grup. Suatu tes mungkin diujikan pada grup random (ekuivalen) dan mungkin diujikan pada grup tidak ekuivalen. Dengan demikian, desain tes butir bersama ini dalam penerapannya dapat dibedakan menjadi dua macam bergantung pada spesifikasi grupnya. Jika kedua grup ekuivalen, maka desain yang digunakan disebut desain tes dengan butir bersama grup ekuivalen, dan jika kedua grup tidak ekuivalen, maka desain yang digunakan disebut desain tes dengan butir bersama grup non ekuivalen. Kedua desain ini dapat diterapkan pada penyetaraan horizontal dan khusus desain tes-anchor grup non ekuivalen dapat juga diterapkan pada penyetaraan vertikal.

Bila desain tes dengan butir bersama hendak digunakan, satu hal yang perlu mendapat perhatian yaitu butir-butir bersama harus mewakili tes total baik mengenai karakteristik, isi, dan statistiknya. Butir-butir bersama merupakan representasi dari butir-butir tes total dan sebaiknya penempatan berada pada posisi yang sama untuk kedua tes ketika tes tersebut diujikan. Desain tes dengan butir bersama adalah desain yang paling menguntungkan, karena kedua grup tidak perlu ekuivalen, sehingga masalah-masalah yang dijumpai dalam dua desain lainnya dapat dieliminasi.

D. Menghubungkan Skor Tes Berdasarkan Teori Tes Klasik

Menurut Hambleton, Swaminathan, dan Rogers (1991 : 123), ada dua jenis metode menghubungkan tes-tes, yakni metode linear dan metode equipercentil. Sedangkan Kollen dan Brennan (2004), selain kedua hal di atas, ada satu metode lagi, yakni metode linear sejajar.

Agar terjadi kekonsistenan, Kollen dan Brennan (1995) menggunakan istilah dan notasi, bentuk X didefinisikan sebagai tes yang baru, bentuk Y merupakan tes yang lama,

x dan y merupakan skor observasi pada X dan Y. Notasi $l_Y(x)$, $pl_Y(x)$, dan $eq_Y(x)$ menyatakan fungsi statistik yang digunakan untuk mentransformasikan skor atas X ke skala skor Y berturut-turut menggunakan metode linear, metode paralel linear, dan metode equipersentil. Misalkan k merupakan indikator untuk grup atau sub populasi, dan misalkan K merupakan nomor total dari grup ($k=1,2,\dots,K$). Sebagai contoh, jika fokusnya jender, $K=2$ dan $k = 1,2$. Jika k tidak dinyatakan, maka persamaan dapat diterapkan untuk keseluruhan populasi.

Dengan menggunakan metode linear, persamaan transformasi untuk keseluruhan populasi dinyatakan dengan :

$$l_Y(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu(X)] + \mu(Y) \dots\dots\dots(7.1)$$

Persamaan transformasi untuk subgrup k dinyatakan dengan

$$l_{Yk}(x) = \frac{\sigma_k(Y)}{\sigma_k(X)} [x - \mu_k(X)] + \mu_k(Y) \dots\dots\dots(7.2)$$

Dengan metode linear sejajar, persamaan transformasi untuk subgrup k dinyatakan dengan

$$pl_{Yk}(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu_k(X)] + \mu_k(Y) \dots\dots\dots(7.3)$$

Adapun perbedaannya dengan metode linear, pada metode linear sejajar untuk keseluruhan populasi berlaku

$$\frac{\sigma(Y)}{\sigma(X)} = \frac{\sigma_k(Y)}{\sigma_k(X)} \dots\dots\dots(7.4)$$

Dengan metode equipersentil, perbedaan isi dan tingkat kesulitan antar tes-tes dideskripsikan dengan transformasi nonlinear. Transformasi didefinisikan dengan menghubungkan skor pada suatu tes X dengan tes Y dengan persamaan :

$$eq_Y(x) = G^{-1}[F(x)] \dots\dots\dots(7.5)$$

dengan F merupakan fungsi distribusi kumulatif dari X, G merupakan fungsi distribusi kumulatif dari Y, dan G^{-1} merupakan invers dari G.

Kesalahan pengukuran standar (*Standard Error, SE*) digunakan untuk mengetahui keakuratan metode-metode menghubungkan skor tes. Menurut Kolen dan Brennan (1995), SE pada desain kelompok ekuivalen ditentukan dengan persamaan :

$$\text{var}[\hat{l}_y(x_i)] \cong \frac{\sigma^2(Y)[1-\rho(X,Y)]}{N} \left\{ 2 + [1 + \rho(X,Y)] \left[\frac{x_i - \mu(X)}{\sigma(X)} \right]^2 \right\} \dots\dots\dots(7.6)$$

Untuk menentukan keakuratan metode, kriteria metode diindikasikan dengan SE yang kecil. Metode menghubungkan skor tes C_1 dikatakan lebih akurat daripada metode menghubungkan skor tes C_2 jika $SE(C_1) < SE(C_2)$.

Contoh penelitian mengenai penyetaraan dengan teori tes klasik misalnya menyetarakan tes dari satu kabupaten ke kabupaten lainnya (Heri Retnawati & Kana Hidayati, 2007). Hasil analisis dari SPSS diperoleh simpangan baku dan rerata dari masing-masing tes, selanjutnya disubstitusikan ke persamaan

$$l_{Yk}(x) = \frac{\sigma_k(Y)}{\sigma_k(X)} [x - \mu_k(X)] + \mu_k(Y) \dots\dots\dots(7.7)$$

merupakan kasus dari metode linear

$$pl_{Yk}(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu_k(X)] + \mu_k(Y)$$

disebabkan oleh

$$\frac{\sigma(Y)}{\sigma(X)} = \frac{\sigma_k(Y)}{\sigma_k(X)}$$

Oleh sebab itu, dalam penelitian ini kedua metode ini dirangkum dalam

satu metode saja, yakni metode linear, yang selanjutnya diperoleh formula penyetaraan dari perangkat tes X ke perangkat tes Y sebagai berikut:

$$p_{Yk}(x) = \frac{4.97191}{4.05135} [x - 18.8744] + 28.1635 \dots\dots\dots(7.8)$$

atau disederhanakan menjadi $p_{Yk}(x) = 1.22722 x + 5.0004$.

Selanjutnya dapat dibuat tabel konversi yang merupakan produk akhir sebagai berikut.

Tabel 7.1. Tabel Konversi dari Tes X ke Tes Y

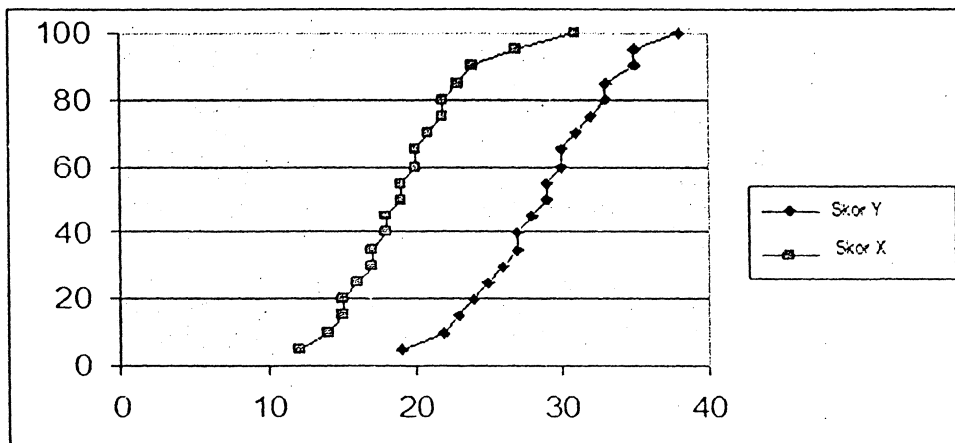
Skor Lama	Skor Baru	Skor Lama	Skor Baru	Skor Lama	Skor Baru
1	6.22762	16	24.63592	31	43.04422
2	7.45484	17	25.86314	32	44.27144
3	8.68206	18	27.09036	33	45.49866
4	9.90928	19	28.31758	34	46.72588
5	11.1365	20	29.5448	35	47.9531
6	12.36372	21	30.77202	36	49.18032
7	13.59094	22	31.99924	37	50.40754
8	14.81816	23	33.22646	38	51.63476
9	16.04538	24	34.45368	39	52.86198
10	17.2726	25	35.6809	40	54.0892
11	18.49982	26	36.90812	41	55.31642
12	19.72704	27	38.13534	42	56.54364
13	20.95426	28	39.36256	43	57.77086
14	22.18148	29	40.58978	44	58.99808
15	23.4087	30	41.817	45	60.2253

Metode penyetaraan lainnya yakni metode Equipercentile. Pada metode ini, persentil pada skor tes dengan perangkat X dan skor tes dengan perangkat Y dihitung terlebih dahulu. Selanjutnya disusun persentil seperti pada Tabel 7.2.

Tabel 7. 2. Persentil Data

Persentil ke-	Skor X	Skor Y	Persentil ke-	Skor X	Skor Y
5	19	12	55	29	19
10	22	14	60	30	20
15	23	15	65	30	20
20	24	15	70	31	21
25	25	16	75	32	22
30	26	17	80	33	22
35	27	17	85	33	23
40	27	18	90	35	24
45	28	18	95	35	27
50	29	19	100	38	31

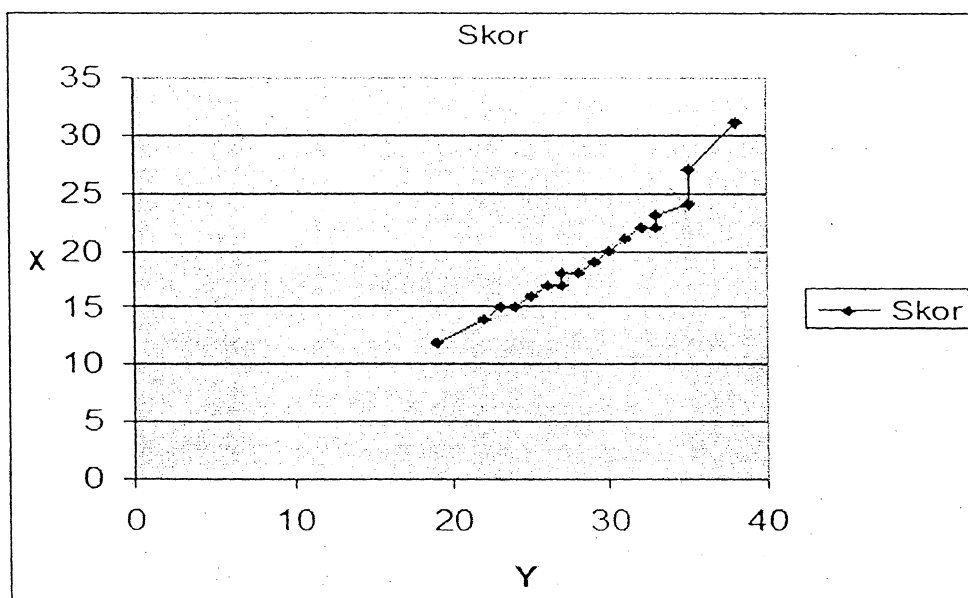
Jika hasil ini disajikan dalam bentuk grafik, maka akan nampak seperti pada Gambar 7.1 berikut.



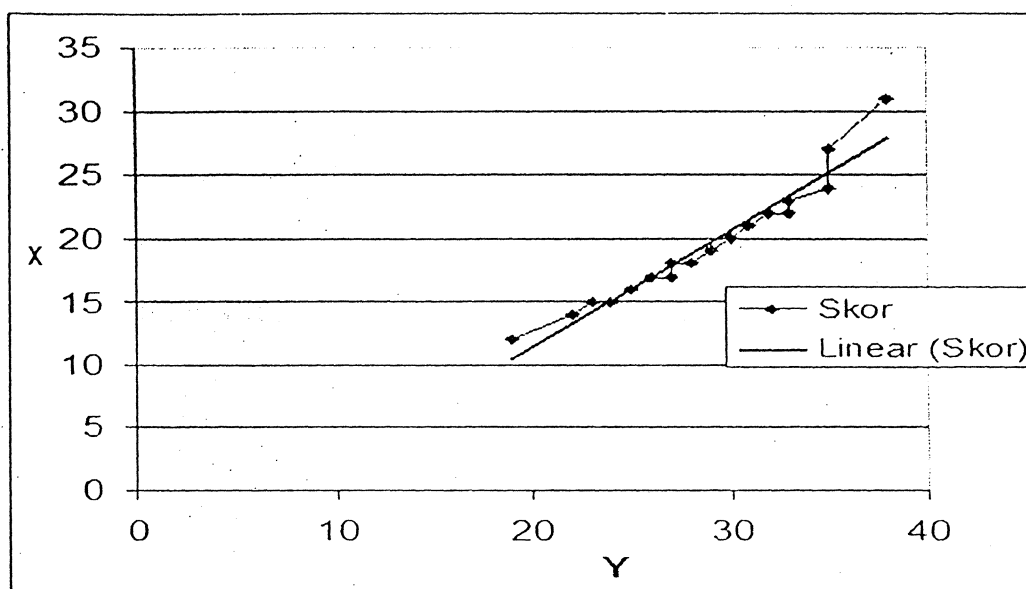
Gambar 7.1. Grafik Persentil Data

Hasil ini menunjukkan bahwa pada persentil yang sama, skor perolehan siswa di kabupaten Y lebih tinggi dibandingkan dengan skor perolehan siswa di kabupaten X.

Untuk menyusun tabel konversi, selanjutnya dibuat grafik hubungan antara skor perolehan siswa di kedua wilayah tersebut Gambar 7.2 dan gambar 7.3 (dengan garis tren) jika akan melakukan prediksi, intrapolasi, atau ekstrapolasi.



Gambar 7.2. Hubungan Antara Skor Perolehan Siswa dengan Tes X dan Skor Siswa dengan Tes Y



Gambar 7.3. Hubungan Antara Skor Perolehan Siswa dengan Tes X dan Dengan Tes Y untuk Ekstrapolasi

Selanjutnya dapat disusun tabel konversi dengan metode equipersentile sebagai berikut.

Tabel 7.3. Tabel Konversi dengan Metode Equipersentile dari Tes X ke Tes Y

Skor Siswa	Skor Siswa yang Baru
12	19
13	20.5
14	22
15	23
16	25
17	26
18	27
19	29
20	30
21	31
22	22
23	33
24	35
25	35.6
26	36.2
27	36.8
28	37.4
29	38

E. Menghubungkan Skor Tes Berdasarkan Teori Respons Butir untuk Data Dikotomi

Dalam teori respons butir, jika model respons butir cocok dengan suatu data set, maka sebarang transformasi linear dari skala pengukuran juga cocok untuk data tersebut. Hal ini berarti, bahwa ada hubungan antara skala pengukuran dari dua tes. Dengan demikian, jika skala tes 1 disetarakan dengan skala tes 2 untuk model 3PL, maka hubungan parameter butir dan kemampuan peserta untuk dua skala tersebut dapat dinyatakan sebagai berikut (Kolen, & Brennan, 1995).

$$\theta_{ii}^* = \alpha \theta_{ii} + \beta, \quad (7.8)$$

$$a_{1j}^* = \frac{a_{1j}}{\alpha} \quad (7.9)$$

$$b_{1j}^* = \alpha b_{1j} + \beta \quad (7.10) \text{ dan}$$

$$c_{1j}^* = c_{1j} \quad (7.11) \text{ dengan}$$

a_{1j} , b_{1j} , dan c_{1j} adalah parameter butir untuk butir j pada skala tes 1,

a_{1j}^* , b_{1j}^* , dan c_{1j}^* adalah parameter butir untuk butir j pada skala tes 1 setelah disetarakan dengan tes 2,

θ_{ii} : kemampuan peserta i pada skala tes 1,

θ_{ii}^* : kemampuan peserta i pada skala tes 1 setelah disetarakan dengan tes 2,

α dan β adalah konstanta penyetaraan.

Parameter tebakan c tidak ditransformasikan karena nilainya tidak bergantung pada metrik θ atau c bebas dari transformasi skala (Kolen & Brennan, 1995). Selanjutnya

konstanta menghubungkan skor tes α dan β dapat dihitung dengan berbagai metode yang disebut metode menghubungkan skor tes.

Terdapat berbagai metode yang dapat diterapkan untuk menghubungkan skor 2 tes atau lebih. Ditinjau dari teknik kalibrasinya, metode menghubungkan skor tes diklasifikasikan menjadi dua metode, yaitu metode kalibrasi terpisah dan metode kalibrasi simultan. Pada metode kalibrasi terpisah, kedua tes dikalibrasi secara sendiri-sendiri, sedangkan pada metode kalibrasi simultan kedua tes dikalibrasi secara serentak atau bersama-sama. Pada kalibrasi simultan, tidak ada perhitungan mengenai konstanta penyetaraan. Hasil kalibrasi kedua tes secara otomatis sudah menunjukkan bahwa parameter butir dan kemampuan sudah berada pada satu skala yang sama.

Pada metode kalibrasi terpisah, ada 2 metode yang termasuk yaitu metode momen dan metode grafik. Pada metode momen, paling sedikit ada tiga metode yang dapat diterapkan, yaitu metode Rerata & Sigma, Rerata & Sigma Tegar, dan Rerata & Rerata. Di antara ketiga metode tersebut, terdapat dua metode yang perlu mendapat perhatian yaitu metode Rerata & Sigma dan Rerata & Rerata. Kedua metode tersebut persamaannya sederhana dan penerapannya sangat mudah.

Pada metode grafik, terdapat dua metode yang dapat diterapkan, yaitu metode kurva karakteristik dari Haebara dan metode kurva karakterististik dari Stocking & Lord. Kedua metode tersebut merupakan metode yang populer dan memberikan hasil penyetaraan yang hampir sama. Pada buku ini dibahas beberapa metode penyetaraan tes antara lain metode rerata dan rerata, rerata dan sigma, Haebara, dan Stocking dan Lord.

1) **Metode Rerata & Rerata**

Pada metode rerata dan rerata, untuk menentukan konstanta penyetaraan melibatkan dua parameter, yaitu parameter daya beda (a) dan tingkat kesukaran butir (b). Konstanta penyetaraan α dan β , dapat dihitung dengan menggunakan rerata dari parameter butir yang terlibat yakni daya beda dan kesukaran butir. Misal penyetaraan dilakukan dari tes 1 ke tes 2, dengan desain tes-anchor menggunakan model 3PL. Hubungan parameter indeks kesukaran dan daya beda adalah sebagai berikut.

$$b_2 = \alpha b_1 + \beta, \quad a_2 = \frac{a_1}{\alpha}$$

Selanjutnya diperoleh

$$\bar{b}_2 = \alpha \bar{b}_1 + \beta,$$

dengan

$$\frac{\bar{a}_2}{a_2} = \frac{\bar{a}_1}{\alpha} \text{ atau } \alpha = \frac{\bar{a}_1}{\bar{a}_2} \dots\dots\dots(7.12)$$

dan

$$\beta = \bar{b}_2 - \alpha \bar{b}_1, \dots\dots\dots(7.13)$$

dengan

\bar{b}_1 dan \bar{b}_2 : berturut-turut rerata indeks kesukaran butir bersama tes 1 dan tes 2,

\bar{a}_1 dan \bar{a}_2 : berturut-turut rerata indeks daya beda butir bersama tes 1 dan tes 2,

α dan β konstanta penyetaraan.

Persamaan (7.12) dan (7.13) digunakan untuk menghitung konstanta penyetaraan tes dengan berdasarkan metode rerata dan rerata.

2) Metode Rerata & Sigma

Pada metode rerata dan sigma, untuk menentukan konstanta penyetaraan α dan β melibatkan rerata dan simpangan baku dari parameter indeks kesulitan yang dapat dijelaskan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991). Misal skor tes 1 disetarakan ke skor tes 2, dengan model 3PL. Berdasarkan model penyetaraannya, hubungan parameter indeks kesulitan butir berhubungan linear,

$$b_2 = \alpha b_1 + \beta,$$

Sehingga diperoleh

$$\bar{b}_2 = \alpha \bar{b}_1 + \beta,$$

dan

$$S_2 = \alpha S_1$$

Jadi

$$\alpha = \frac{S_2}{S_1}, \quad \dots\dots\dots(7.14)$$

dan

$$\beta = \bar{b}_2 - \alpha \bar{b}_1, \quad \dots\dots\dots(7.15)$$

dengan

\bar{b}_1 dan \bar{b}_2 : berturut-turut rerata indeks kesukaran butir tes 1 dan tes 2,

S_1 dan S_2 : berturut-turut simpangan baku indeks kesukaran butir tes 1 dan tes 2,

α dan β konstanta penyetaraan.

3) Metode Haebara

Dalam menentukan konstanta penyetaraan pada metode rerata dan sigma maupun rerata dan rerata tidak melibatkan semua parameter butirnya secara simultan. Terdapat suatu metode penyetaraan yang melibatkan semua parameter butirnya secara simultan, yaitu metode kurva karakteristik yang diusulkan oleh Haebara dan metode kurva karakteristik dikembangkan oleh Stocking & Lord. Metode ini mempunyai 2 jenis rumus dalam penerapannya.

Jenis yang pertama yaitu konstanta penyetaraan dihitung dengan cara menentukan selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteristik butir dari dua skala yang sudah disetarakan, dikuadratkan kemudian dijumlahkan. Jenis yang kedua yaitu konstanta penyetaraan dihitung dengan cara menentukan selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteristik tes dari dua skala yang sudah disetarakan lalu dikuadratkan. Selanjutnya dengan menggunakan kriteria tertentu, konstanta penyetaraan α dan β diperoleh dengan meminimumkan fungsi tertentu yang memuat variabel α dan β .

Metode Haebara adalah metode kurva karakteristik yang penyetaraan parameter butirnya berdasarkan pada fungsi karakteristik butir. Prosedur

komputasinya menggunakan variasi yang pertama, yang dapat dijelaskan sebagai berikut (Kolen, & Brennan, 1995). Jumlah kuadrat dari selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteristik butir dari dua skala yang sudah disetarakan dinyatakan dengan $H(\theta_i)$ yaitu

$$H(\theta_i) = \sum_{j=1}^n (T_{ij} - T_{ij}^*)^2 \dots\dots\dots(7.16)$$

dengan

$$T_{ij} = P_j(\theta_i),$$

$$T_{ij}^* = P_j^*(\theta_i),$$

dan

n : banyaknya butir anchor,

$P_j(\theta_i)$: Probabilitas menjawab benar butir j oleh peserta berkemampuan θ_i ,

$P_j^*(\theta_i)$: Probabilitas hasil transformasinya.

serta transformasi pada butir anchor,

$$b_j^* = \alpha b_j + \beta, \quad a_j^* = \frac{a_j}{\alpha}, \quad \text{dan} \quad c_j^* = c_j.$$

Didefinisikan fungsi yang persamaannya sebagai berikut:

$$F = \frac{1}{N} \sum_{i=1}^N H(\theta_i) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n (T_{ij} - T_{ij}^*)^2 \dots\dots\dots(7.17)$$

dengan N sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Fungsi F pada persamaan (7.17) merupakan fungsi dalam α dan β . Selanjutnya konstanta penyetaraan α dan β dipilih sedemikian rupa sehingga fungsi F minimum. Fungsi F mencapai nilai minimum bila

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0 \dots\dots\dots(7.18)$$

Persamaan (7.18) non linear dan mempunyai solusi numerik, sehingga persamaan tersebut dapat diselesaikan dengan menggunakan prosedur numerik. Salah satu metode yang dapat digunakan untuk menyelesaikan persamaan tersebut adalah metode numerik Newton Raphson.

4) Metode Stocking & Lord

Metode Stocking dan Lord adalah metode kurva karakteristik yang penyetaraan parameter butirnya berdasarkan pada fungsi karakteristik tes. Formula komputasinya menggunakan variasi yang kedua, prosedur komputasinya disajikan sebagai berikut (Kolen, & Brennan, 1995: 170). Kuadrat dari selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteris tes dari dua skala yang sudah disetarakan dinyatakan dengan $SL(\theta_i)$ yaitu

$$SL(\theta_i) = (T_i - T_i^*)^2 \dots\dots\dots(7.19)$$

dengan

$$T_i = \sum_{j=1}^n P_j(\theta_i)$$

$$T_i^* = \sum_{i=1}^n P_j^*(\theta_i)$$

dan

n : panjang tes-anchor,

$P_j(\theta_i)$: Probabilitas menjawab benar butir j oleh peserta berkemampuan θ_i ,

$P_j^*(\theta_i)$: Probabilitas hasil transformasinya.

T_i : Skor murni peserta berkemampuan θ_i pada tes dasar,

T_i^* : Skor murni hasil tranformasinya.

Dengan transformasi pada tes dengan butir bersama,

$$b_j^* = \alpha b_j + \beta, \quad a_j^* = \frac{a_j}{\alpha}, \quad \text{dan} \quad c_j^* = c_j.$$

Di definisikan fungsi

$$F = \frac{1}{N} \sum_{i=1}^N (T_i - T_i^*)^2 \dots\dots\dots(7.20)$$

dengan N adalah sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Selanjutnya konstanta penyetaraan α dan β dipilih sehingga fungsi F minimum. Fungsi F pada persamaan (7.20), mencapai minimum bila

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0 \dots\dots\dots(7.21)$$

Persamaan (7.21) non linear dan mempunyai solusi numerik, sehingga persamaan tersebut hanya dapat diselesaikan dengan menggunakan prosedur numerik. Salah satu metode yang dapat digunakan untuk menyelesaikan persamaan tersebut adalah metode numerik Newton Raphson.

Berikut disajikan contoh analisis equating dengan metode rerata dan sigma pada 9 paket ujian bahasa Inggris dengan model 1PL. Kedelapan perangkat lain disetarakan ke perangkat 2. Respons peserta tiap paket dikaliberasi secara terpisah, kemudian tingkat kesulitannya disetarakan ke paket 2. Pada tiap tiket dihitung rerata dan standar deviasi tingkat kesulitannya.

Dengan metode rerata dan sigma, untuk menentukan konstanta penyetaraan α dan β melibatkan rerata dan simpangan baku dari parameter indeks kesulitan. Misal skor tes paket 3 disetarakan dengan skor tes 2, hubungan parameter indeks kesulitan butir berhubungan linear,

$$b_2 = \alpha b_3 + \beta,$$

Sehingga diperoleh

$$\bar{b}_2 = \alpha \bar{b}_3 + \beta,$$

dan

$$S_2 = \alpha S_3$$

Jadi

$$\alpha = \frac{S_2}{S_3},$$

dan

$$\beta = \bar{b}_2 - \alpha \bar{b}_3,$$

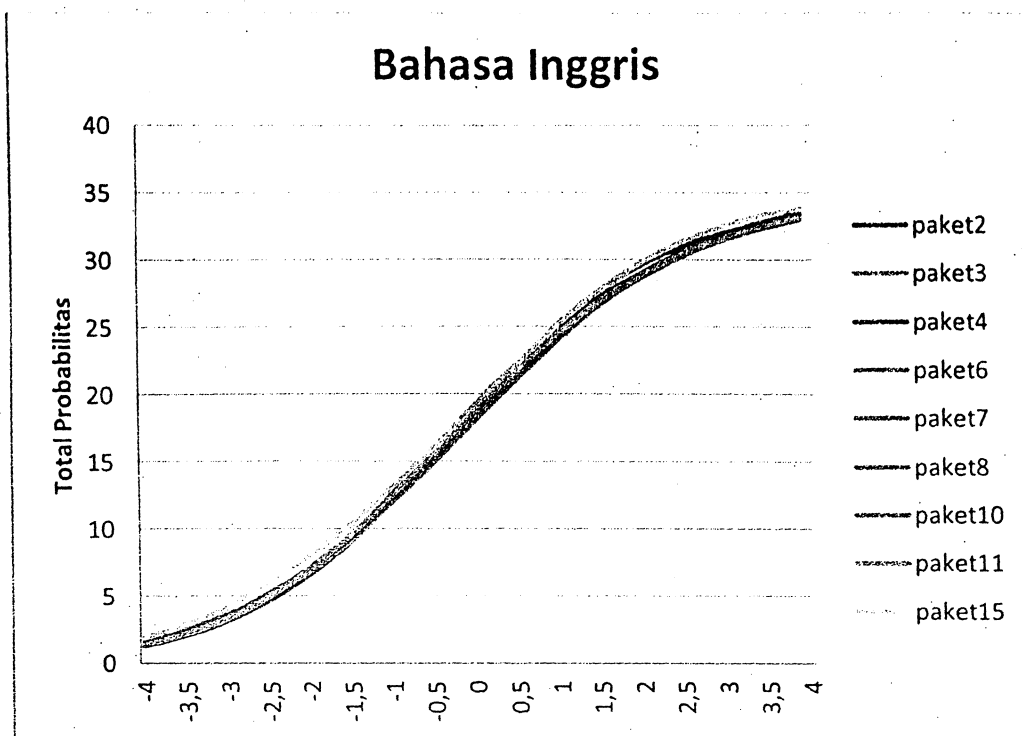
dengan

\bar{b}_3 dan \bar{b}_2 : berturut-turut rerata indeks kesukaran butir tes 3 dan tes 2,

S_3 dan S_2 : berturut-turut simpangan baku indeks kesukaran butir tes 3 dan tes 2,

α dan β konstanta penyetaraan.

Hasil analisis pada studi ini berupa tingkat kesulitan butir yang telah disetarakan ke paket 2 (induk) pada setiap mata pelajaran yang diujikan. Selain dihasilkan tabel konversi, tingkat kesulitan hasil penyetaraan beberapa paket ini kemudian dimanfaatkan untuk membuat kurva karakteristik tes (*test characteristics curve*, TCC) berdasarkan teori respons butir untuk setiap paket. Hasil TCC untuk tiap paket dibandingkan dengan paket lain. Suatu paket, misalkan paket-m dikatakan setara dengan paket 2, ditunjukkan dengan kedekatan TCC dari paket-m dengan TCC dari paket 2. Semakin dekat kedua kurva tersebut, maka kedua perangkat tersebut dikatakan setara. Hal ini terjadi karena perangkat tes yang digunakan memberikan total probabilitas menjawab benar yang hampir sama pada setiap skala kemampuan siswa, seperti disajikan pada Gambar 7.4.



Gambar 7.4. Penyetaraan 9 Perangkat Bahasa Inggris

F. Metode Penyetaraan Pada IRT dengan Data Politomi

Dalam teori respons butir, jika suatu data set telah cocok dianalisis dengan model respons butir, maka sebarang transformasi linear dari skala pengukuran juga cocok untuk data tersebut. Pada data politomi hal ini berarti bahwa ada hubungan linear antara skala pengukuran dari dua tes. Dengan demikian, jika skala tes 1 disetarakan dengan skala tes 2 untuk model penskoran politomi, maka hubungan parameter butir dan kemampuan peserta untuk dua skala tersebut dapat dinyatakan sebagai berikut (Kolen, & Brennan, 1995).

$$\theta_{ii}^* = \alpha \theta_{ii} + \beta,$$

$$a_{ij}^* = \frac{a_{ij}}{\alpha}$$

$$b_{ij}^* = \alpha b_{ij} + \beta$$

dan

$$c_{1j}^* = c_{1j}$$

dengan

a_{ij} , b_{ij} , dan c_{ij} adalah parameter butir untuk butir j pada skala tes 1,

a_{ij}^* , b_{ij}^* , dan c_{ij}^* adalah parameter butir untuk butir j pada skala tes 1 setelah disetarakan dengan tes 2,

θ_{ii} : kemampuan peserta i pada skala tes 1,

θ_{ii}^* : kemampuan peserta i pada skala tes 1 setelah disetarakan dengan tes 2,

α dan β adalah konstanta penyetaraan.

Parameter tebakan c tidak ditransformasikan karena nilainya tidak bergantung pada metrik θ atau c bebas dari transformasi skala (Kolen & Brennan, 1995: 163). Selanjutnya konstanta penyetaraan α dan β dapat dihitung dengan berbagai metode yang disebut metode penyetaraan. Dalam teori respons butir, metode penyetaraan yang diterapkan berasumsi bahwa kedua tes yang disetarakan unidimensi.

Seperti halnya pada dikotomi, pada politomi pun terdapat berbagai metode yang dapat diterapkan untuk menyetarakan tes. Ditinjau dari teknik kalibrasinya, metode penyetaraan tes diklasifikasikan menjadi dua metode, yaitu metode kalibrasi terpisah dan metode kalibrasi simultan. Pada kalibrasi simultan, tidak ada komputasi mengenai konstanta penyetaraan, hasil kalibrasi kedua tes secara otomatis sudah berada pada satu skala. Dalam metode kalibrasi terpisah, yaitu metode momen dan metode grafik. Pada metode momen, paling sedikit ada tiga metode yang dapat diterapkan, yaitu metode Rerata & Sigma, Rerata & Sigma Tegar, dan Rerata & Rerata. Diantara ketiga metode tersebut, terdapat dua metode yang perlu mendapat perhatian yaitu metode Rerata & Sigma dan Rerata & Rerata. Pada metode grafik, terdapat dua metode yang dapat diterapkan, yaitu metode kurva karakteristik dari Haebara dan metode kurva karakteristik dari Stocking & Lord. Pada buku ini hanya disajikan model GPCM, dan untuk model lainnya dapat menyesuaikan persamaan.

1) Metode Rerata & Rerata

Seperti halnya metode rerata & rerata pada penyetaraan tes yang menggunakan model dikotomus (3PL), dalam menentukan konstanta penyetaraan dengan metode rerata dan rerata pada penyetaraan tes yang menggunakan GPCM, juga melibatkan dua parameter, yaitu parameter daya beda dan kesukaran butir. Nilai α dihitung berdasarkan rerata daya beda butir-butir bersama (common item) dari masing-masing grup, sedangkan nilai β didapat dengan menghitung rerata dari semua parameter kesukaran butir bersama (*anchor item*). Perhitungan nilai β sama dengan perhitungan nilai β pada metode rerata dan sigma. Rerata dari semua parameter kategori n butir anchor untuk grup g adalah sebagai berikut (Nony Swediati, 1997: 38).

$$\bar{b}_g = \frac{1}{mn} \sum_{j=1}^n \sum_{k=1}^m b_{jkg}, (g=1,2) \dots\dots\dots(7.22)$$

Selanjutnya konstanta penyetaraan didapat, yaitu:

$$\alpha = \frac{\bar{a}_1}{\bar{a}_2} \dots\dots\dots(7.23)$$

dan

$$\beta = \bar{b}_1 - \alpha \bar{b}_2 \dots\dots\dots(7.24)$$

Jadi untuk menghitung konstanta penyetaraan dengan metode Rerata & Rerata pada penyetaraan tes yang menggunakan GPCM berdasarkan persamaan (7.23) dan (7.24).

Metode Rerata & Rerata memiliki kelebihan dan kekurangan. Kelebihan metode tersebut pada penyetaraan tes yang menggunakan GPCM adalah formulanya sederhana, mudah diterapkan, dan komputasinya dapat dilakukan dengan cara manual atau program komputer yang dapat dikembangkan sendiri. Dapat pula cara ini diterapkan pada model penskoran campuran secara simultan, misalnya penskoran dikotomi dan politomi. Kekurangan metode Rerata dan Rerata

yaitu parameter butir diperlakukan sebagai satuan (*entity*) yang bebas dan pencilan (*outliers*) berpengaruh pada hasil penyetaraan tes.

Setelah α dan β dihitung pada model GPCM, kemudian nilai parameter butir dan kemampuan dari tes 1 ditempatkan pada skala yang sama dengan skala dari tes 2 menggunakan hubungan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991: 129).

$$b_2^* = \alpha b_1 + \beta \dots\dots\dots(7.25)$$

$$a_2^* = \frac{a_1}{\alpha} \dots\dots\dots(7.26)$$

$$\theta_2^* = \alpha \theta_1 + \beta \dots\dots\dots(7.27)$$

dengan

b_2^* : indeks kesukaran butir-butir dalam tes1 ditempatkan pada skala tes 2,

a_2^* : indeks daya beda butir-butir dalam tes 1 ditempatkan pada skala tes 2.

θ_2^* : kemampuan para peserta tes 1 ditempatkan pada skala tes 2.

Persamaan (7.25), (7,26), dan (7.27) merupakan rumus konversi penyetaraan tes.

2) Metode Rerata & Sigma pada GPCM

Prosedur metode Rerata dan Sigma pada penyetaraan tes yang menggunakan GPCM relatif sama dengan prosedur metode Rerata dan Sigma pada penyetaraan tes yang menggunakan model dikotomus 3PL. Dalam proses penghitungan penyetaraannya, metode ini melibatkan rerata dan simpangan baku parameter kategori butir. Indeks kesukaran butir untuk setiap butir pada model politomi, terdiri dari beberapa parameter kategori butir. Nilai konstanta penyetaraan didapat dengan menghitung rerata dan simpangan baku dari semua parameter kategori butir bersama.

Rerata dan simpangan baku dari semua parameter kategori n butir bersama dengan m parameter kategori untuk masing-masing grup adalah sebagai berikut (Nony Swediati, 1997: 38).

$$\bar{b}_1 = \frac{1}{mn} \sum_{j=1}^n \sum_{k=1}^m b_{jk1} = \text{rata-rata } b_{jk1} \dots\dots\dots(7.28)$$

$$\bar{b}_2 = \frac{1}{mn} \sum_{j=1}^n \sum_{k=1}^m b_{jk2} = \text{rata-rata } b_{jk2} \dots\dots\dots(7.29)$$

$$S_{b1} = \left\{ \sum_{j=1}^n \sum_{k=1}^m \frac{(b_{jk1} - \bar{b}_1)^2}{mn-1} \right\}^{\frac{1}{2}} = \text{simpangan baku } b_{jk1} \dots\dots\dots(7.30)$$

$$S_{b2} = \left\{ \sum_{j=1}^n \sum_{k=1}^m \frac{(b_{jk2} - \bar{b}_2)^2}{mn-1} \right\}^{\frac{1}{2}} = \text{simpangan baku dari } b_{jk2} \dots\dots\dots(7.31)$$

Selanjutnya diperoleh konstanta penyetaraan sebagai berikut.

$$\alpha = \frac{S_{b2}}{S_{b1}} \dots\dots\dots(7.32)$$

dan

$$\beta = \bar{b}_2 - \alpha \bar{b}_1 \dots\dots\dots(7.33)$$

dengan

b_{jk1} : indeks kesukaran kategori k butir j untuk tes1,

b_{jk2} : indeks kesukaran kategori k butir j untuk tes 2,

a_{j1} : indeks daya beda butir j untuk tes1,

a_{j2} : indeks daya beda butir j untuk tes2,

n : banyaknya butir *anchor*, dan

m+1 : banyaknya kategori butir politomi.

Persamaan (7.32) dan (7.33) digunakan untuk menghitung konstanta penyetaraan tes yang menggunakan GPCM dengan metode Rerata dan Sigma. Setelah α dan β dihitung, baik untuk model dikotomi maupun politomi, kemudian nilai parameter butir dan kemampuan dari tes 1 ditempatkan pada skala yang sama dengan skala dari tes 2 menggunakan rumus konversi.

3) Metode Haebara pada GPCM

Metode Haebara untuk model dikotomus 3PL dapat diperluas untuk model politomi (GPCM). Prinsipnya sama, prosedur dan langkah-langkahnya sama, hanya berbeda dalam definisi skor murni (*true score*) peserta. Berikut disajikan prosedur penerapan metode Haebara untuk model GPCM.

Jumlah kuadrat dari selisih antara nilai fungsi untuk absis yang sama pada masing-masing kurva karakteristik butir dari dua skala yang sudah disetarakan adalah:

$$H(\theta_i) = \sum_{j=1}^n (T_{ij} - T_{ij}^*)^2,$$

dengan

$$T_{ij} = \sum_{k=1}^{m-1} P_{jk}(\theta_i) \dots\dots\dots(7.34)$$

$$T_{ij}^* = \sum_{k=1}^{m-1} P_{jk}^*(\theta_i) \dots\dots\dots(7.35)$$

dan

k : kategori respons butir dari butir j,

m : banyaknya kategori butir j,

n : banyaknya butir-butir-butir bersama,

$P_{jk}(\theta_i)$: probabilitas peserta berkemampuan θ_i memperoleh skor kategori k pada butir j,

$P_{jk}^*(\theta_i)$: probabilitas hasil transformasinya,

serta

$$P_{jk}(\theta_i) = \frac{\exp\left[\sum_{v=1}^k Da_j(\theta_i - b_{jv})\right]}{\sum_{c=1}^m \exp\left[\sum_{v=1}^c Da_j(\theta_i - b_{jv})\right]}, \quad k = 0, 1, 2, \dots, m$$

serta transformasi pada butir-butir bersama,

$$\overline{b_{jk}^*} = \alpha \overline{b_{jk}} + \beta, \quad a_j^* = \frac{a_j}{\alpha},$$

Kemudian didefinisikan fungsi

$$F = \frac{1}{N} \sum_{i=1}^N H(\theta_i) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n (T_{ij} - T_{ij}^*)^2 \dots\dots\dots(7.36)$$

dengan N adalah sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Fungsi F merupakan fungsi dalam α dan β . Selanjutnya konstanta penyetaraan α dan β diperoleh dengan cara meminimkan nilai fungsi F pada persamaan (7.36). Teknik komputasinya sama dengan teknik komputasi yang diterapkan untuk model 3PL. Setelah konstanta penyetaraan α dan β didapat, nilai parameter butir dan kemampuan dari skala tes 1 ditempatkan pada skala tes 2 dengan menggunakan rumus konversi pada persamaan.

4) Metode Stocking & Lord pada GPCM

Metode kurva karakteristik untuk model dikotomus dapat diperluas untuk model politomi. Baker (Nony Swediati, 1997: 29) telah memperluas metode kurva karakteristik dari Stocking & Lord model dikotomus ke model politomi GRM. Prinsipnya sama, prosedur dan langkah-langkahnya sama, hanya berbeda dalam

definisi skor murni (true score) peserta. Berikut prosedur penerapan metode kurva karakteristik pada penyetaraan tes yang menggunakan GPCM.

Mula-mula didefinisikan skor murni peserta berkemampuan θ_i pada tes dasar dan skor murni hasil transformasinya sebagai berikut.

$$T_i = \sum_{j=1}^n \sum_{k=1}^{m-1} kP_{jk}(\theta_i) \dots\dots\dots(7.37)$$

$$T_i^* = \sum_{j=1}^n \sum_{k=1}^{m-1} kP_{jk}^*(\theta_i) \dots\dots\dots(7.38)$$

dengan

k : kategori respons butir dari butir j,

m : banyaknya kategori butir j,

n : banyaknya butir tes-anchor,

$P_{jk}(\theta_i)$: probabilitas peserta berkemampuan θ_i memperoleh skor kategori pada butir j,

$P_{jk}^*(\theta_i)$: probabilitas hasil transformasinya,

dan

$$P_{jk}(\theta_i) = \frac{\exp\left[\sum_{v=1}^k Da_j(\theta_i - b_{jv})\right]}{\sum_{c=1}^m \exp\left[\sum_{v=1}^c Da_j(\theta_i - b_{jv})\right]}, k = 0, 1, 2, \dots, m$$

serta transformasi pada tes-anchor,

$$\overline{b_{jk}^*} = \alpha \overline{b_{jk}} + \beta, a_j^* = \frac{a_j}{\alpha},$$

Kemudian didefinisikan fungsi

$$F = \frac{1}{N} \sum_{i=1}^N (T_i - T_i^*)^2 \dots\dots\dots(7.39)$$

dengan N adalah sebarang bilangan asli menyatakan banyaknya titik pada skala θ . Fungsi F merupakan fungsi dalam α dan β . Selanjutnya konstanta penyetaraan α dan β diperoleh dengan cara meminimkan nilai fungsi F pada persamaan (7.39). Teknik komputasi metode Stocking & Lord pada penyetaraan tes yang menggunakan GPCM sama dengan teknik komputasi yang diterapkan pada penyetaraan tes yang menggunakan model dikotomis 3PL.

Kelebihan dan kekurangan dari metode Stocking & Lord pada model politomi sama dengan penerapan metode ini pada data dikotomi. Kelemahan lain yakni model ini tidak dapat diterapkan pada model penskoran campuran secara bersama-sama (simultan).

Contoh analisis pada data politomi yakni analisis penyetaraan skor dengan butir soal Indonesia dengan skor dengan butir soal Internasional pada data TIMSS 2007 mata pelajaran matematika. Ada 19 butir soal TIMSS mata pelajaran matematika yang diskor dengan 3 kategori (0 salah, 1 betul sebagian, 2 betul seluruhnya).

Dengan menggunakan data Indonesia dan data Internasional, parameter butir diestimasi terpisah. Hasil estimasi parameter butir disajikan pada Tabel 7.1.

Tabel 7.1. Hasil Estimasi Parameter Butir (Indonesia dan Internasional)

No.	Kode	Materi	Jabaran Materi	Tingkat Berfikir	Indonesia (Sebelum Equating)			Internasional			Indonesia (setelah Equating)		
					Slope	Step 1	Step 2	Slope	Step 1	Step 2	Slope	Step 1	Step 2
					a	b1	b2	a	b1	b2	a	b1	b2
1	M022232	Number	Fractions and Decimals	Applying	0,98	-3,06	-0,88	0,53	-3,77	0,59	0,916	-2,615	-0,283
2	M022234A	Geometry	Geometric Shapes	Applying	1,2	-1,38	-0,74	0,8	-1,4	-0,14	1,122	-0,818	-0,133
3	M022234B	Number	Ratio, Proportion and Percent	Applying	1,3	-2,22	-0,38	0,9	-2,56	0,4	1,215	-1,716	0,2522
4	M042220	Data and Chance	Data Organization and Representation	Applying	0,79	-1,91	0,15	0,69	-1,96	1,08	0,738	-1,385	0,8192
5	M042304B	Number	Ratio, Proportion and Percent	Applying	1,44	-0,89	-1,41	0,97	-0,26	-1,08	1,346	-0,293	-0,85
6	M042304D	Number	Whole Numbers	Reasoning	0,63	-1,97	0,01	0,58	-1,64	1,18	0,589	-1,449	0,6694
7	M042303B	Data and Chance	Data Interpretation	Reasoning	0,47	-1,47	-2,33	0,37	-0,84	-0,74	0,439	-0,914	-1,834
8	M032640	Algebra	Patterns	Reasoning	0,4	-3,18	-1,06	0,61	-2,36	-0,76	0,374	-2,744	-0,475
9	M032755	Number	Ratio, Proportion and Percent	Reasoning	0,91	-2,18	-1,66	1,1	-1,51	-0,93	0,851	-1,674	-1,117
10	M032753A	Data and Chance	Data Interpretation	Reasoning	0,78	-1,68	-1,72	1,12	-1,03	-0,45	0,729	-1,139	-1,182
11	M032753B	Data and Chance	Data Interpretation	Reasoning	0,94	-1,54	-2,76	1,21	-0,94	-0,88	0,879	-0,989	-2,294
12	M042059	Number	Ratio, Proportion and Percent	Knowing	0,81	-0,57	-0,63	0,77	-0,26	0,1	0,757	0,0489	-0,015
13	M042207	Data and Chance	Data Organization and Representation	Applying	0,46	-2,54	1,46	0,44	-2,97	3,01	0,43	-2,059	2,2208
14	M032695	Data and Chance	Data Organization and Representation	Applying	0,62	-0,86	0,38	0,55	-0,86	1,26	0,579	-0,261	1,0653
15	M032683	Algebra	Algebraic Expression	Knowing	0,79	-1,96	-0,02	0,49	-2,47	0,73	0,738	-1,438	0,6373
16	M032757	Algebra	Patterns	Reasoning	0,49	-0,67	1,95	0,48	-2,08	2,46	0,458	-0,058	2,7451
17	M032760A	Algebra	Patterns	Reasoning	0,7	-1,67	0,05	0,81	-2,06	0,72	0,654	-1,128	0,7122
18	M032761	Algebra	Algebraic Expression	Reasoning	0,89	-2,09	-1,67	1,05	-1,66	-0,84	0,832	-1,577	-1,128
19	M032692	Geometry	Geometric Shapes	Reasoning	0,55	-4,95	-0,47	0,69	-1,97	0,01	0,514	-4,637	0,1559

Kemudian diestimasi rerata parameter a dan b pada kedua kelompok, hasilnya disajikan pada tabel 7.2. dan selanjutnya diestimasi α dan β .

Tabel 7.2. Rerata parameter a dan b

Rerata	Indonesia	Internasional
a	0.797368	0.745263
b	-1.27684	-0.70737

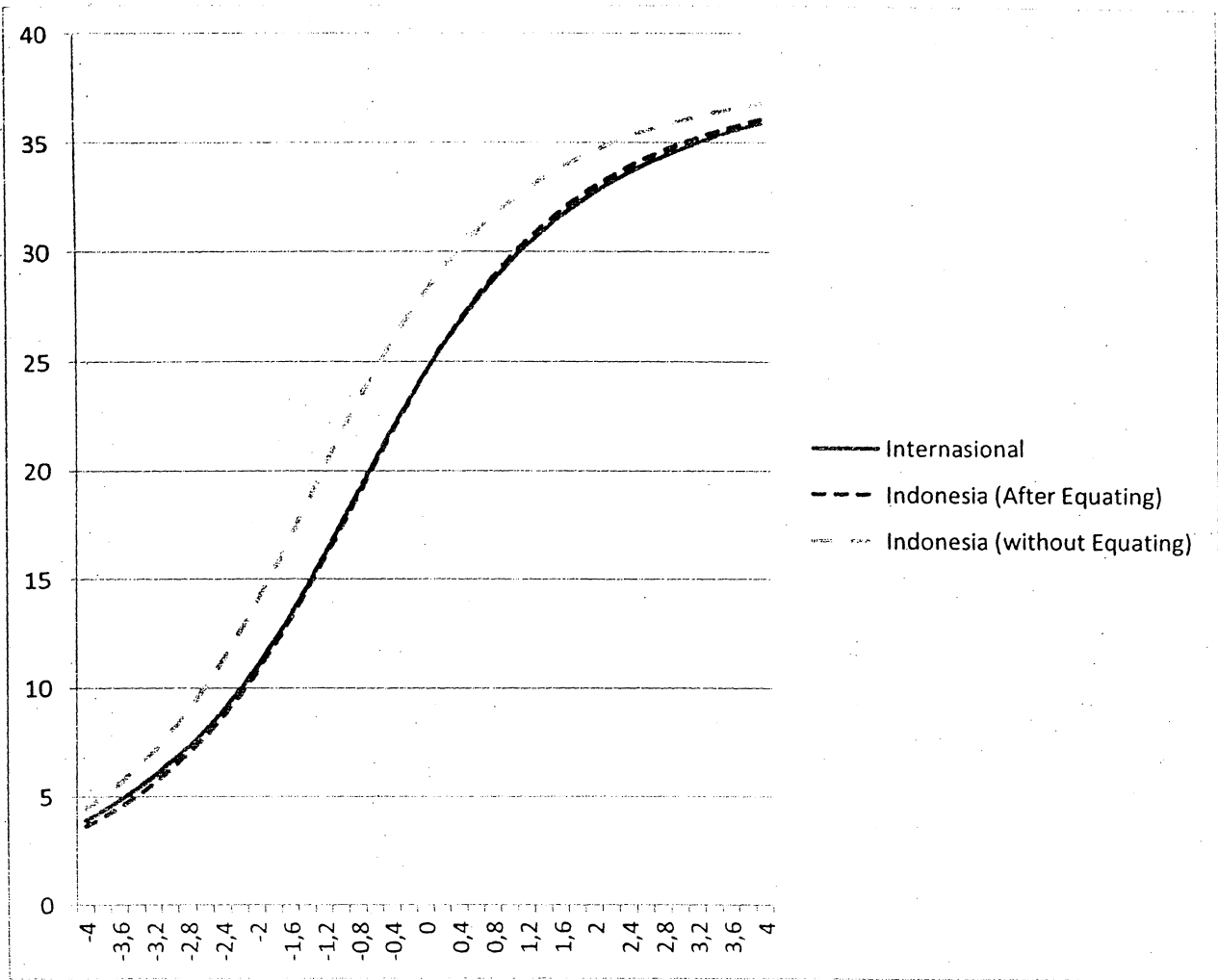
$$\alpha = 1,069915$$

$$\beta = 0,658744$$

Sehingga diperoleh persamaan untuk penyetaraan

$$b_2^* = 1,0699b_1 + 0,6587 \quad a_2^* = \frac{a_1}{1,0699} \quad \text{dan} \quad \theta_2^* = 1,0699 \theta_1 + 0,6587.$$

Persamaan tersebut digunakan untuk transformasi parameter butir yang disajikan pada Tabel 7.1 kolom setelah penyetaraan. Hasil tersebut kemudian digunakan untuk menggambar kurva karakteristik tes dari 19 butir politomi, yang disajikan pada Gambar 7.3.



Gambar 7.3. Kurva Karakteristik Tes Indonesia dan Internasional.

Dengan cara yang sama dengan model GPCM, dapat dikembangkan persamaan untuk penyetaraan pada model lainnya, misal GRM, PCM, maupun NRM.

* * *

Daftar Pustaka

- Brennan, R.L., dan Kolen, M.J. (2004). *Concordance Between ACT and ITED Scores From Different Population*. *Jurnal Applied Psychological Measurement*, Vol 28. No. 4, July 2004, p. 219-226
- Croker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.
- Dorans, N.J. (2004). *Equating, Concordance and Expectation*. *Jurnal Applied Psychological Measurement*, Vol 28. No. 4, July 2004, p. 219-226
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Heri Retnawati, dkk. Sedang dalam proses publikasi. *Equating in Politomus Data (Studi pada Data TIMSS 2007 Mata Pelajaran Matematika)*. *Paper*.
- Heri Retnawati & Kana Hidayati. (2007). *Perbandingan metode concordance berdasarkan teori tes klasik*. *Laporan penelitian*. Lembaga Penelitian UNY Yogyakarta.
- Kolen, M.J. (2004). *Linking Assesment : Concept and History*. *Jurnal Applied Psychological Measurement*, Vol 28. No. 4, July 2004, p. 219-226.
- Kolen, M.J. dan Brennan, R.L. (2004). *Test Equating : Methods and Practices*. New York : Springer.
- Nonny Swediati. (1997). *Equating tests under the Generalized Partial Credit Model*. *Doctoral Dissertations*. University of Massachusetts at Amherst.