

*Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA,  
Fakultas MIPA, Universitas Negeri Yogyakarta, 14 Mei 2011*

**MENGESTIMASI KEMAMPUAN PESERTA TES URAIAN MATEMATIKA  
DENGAN PENDEKATAN TEORI RESPONS BUTIR  
DENGAN PENSKORAN POLITOMUS  
DENGAN *GENERALIZED PARTIAL CREDIT MODEL***

**Heri Retnawati**

*Jurusan Pendidikan Matematika, FMIPA, Universitas Negeri Yogyakarta  
Email: retnawati\_heriuny@yahoo.co.id*

**Abstrak**

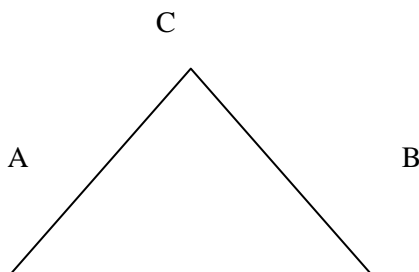
Dalam pelaksanaan tes uraian, penskoran biasanya dilakukan secara parsial berdasarkan langkah-langkah yang harus ditempuh untuk menjawab benar suatu butir soal. Penskoran dilakukan perlangkah dan skor perbutir diperoleh peserta diperoleh dengan menjumlah skor siswa tiap langkah, dan kemampuan diestimasi dengan skor mentah. Model penskoran seperti ini belum tentu tepat, karena tingkat kesulitan tiap langkah tidak diperhitungkan. Pendekatan alternatif yang dapat digunakan yakni pendekatan teori respons butir untuk penskoran politomus, salah satunya dengan *generalized partial credit model* (GPCM). Pada artikel ini disajikan penskoran butir pada GPCM, model matematikanya, estimasi parameternya, dan estimasi kemampuan peserta disertai kasus dan pemanfaatan *software* Parscale dari SSI.

**Kata kunci:** teori respons butir, *generalized partial credit model*

**PENDAHULUAN**

Ada berbagai bentuk tes yang biasa digunakan dalam pendidikan ataupun psikologi (Gronlund & Linn, 1990). Demikian pula bentuk tes dalam pendidikan matematika. Berbagai bentuk tes yang sering digunakan dalam pendidikan matematika diantaranya bentuk objektif dan uraian. Penskoran bentuk objektif biasanya dengan model dikotomi, betul diskor 1 dan salah diskor 0. 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 1 sebagai berikut.

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

Pada pendekatan teori tes klasik, perhitungan skor dilakukan dengan menjumlah skor keseluruhan yang diperoleh siswa. Pendekatan ini belum tentu tepat, karena tingkat kesulitan tiap langkah tidak diperhitungkan. Pendekatan alternatif yang dapat digunakan yakni pendekatan teori respons butir untuk penskoran politomus, salah satunya dengan *generalized partial credit model* (GPCM) yang merupakan perluasan dari *Partial Credit Model* (PCM) (Muraki, 1999). Pada GPCM, tingkat kesulitan tiap langkah diperhitungkan untuk mengestimasi kemampuan peserta. Penskoran tidak dilakukan secara langsung, namun melalui menggunakan metode tertentu setelah estimasi parameter butir dilakukan. Pada artikel ini disajikan penskoran butir pada GPCM, model matematikanya, estimasi parameternya, dan estimasi kemampuan peserta disertai kasus dan pemanfaatan *software* Parscale dari SSI.

**PEMBAHASAN**

Penskoran dengan menggunakan skor menjawab benar pada jawaban siswa dengan model politomus tidak sepenuhnya tepat. Hal ini disebabkan karena tingkat kesukaran tiap langkah tidak diperhitungkan. Selain itu, peluang menjawab benar seorang siswa berdasarkan respons tertentu tidak dapat diprediksikan (Keeves & Alagumalai, 1999). Terkait dengan hal tersebut, perlunya pendekatan lain, diantaranya menggunakan teori respons butir.

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.

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

Dengan

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

$\theta$  : kemampuan peserta,

$m+1$  : banyaknya kategori butir j,

$b_{jk}$  : indeks kesukaran kategori k butir j

$$\sum_{h=0}^k (\theta - b_{jh}) \equiv 0 \quad \text{dan} \quad \sum_{h=0}^h (\theta - b_{jh}) \equiv \sum_{h=1}^h (\theta - b_{jh}) \dots\dots\dots(2)$$

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 (1997) merupakan bentuk umum dari PCM, yang dinyatakan dalam bentuk matematis, yang disebut sebagai fungsi respons kategori butir sebagai berikut.

$$P_{jh}(\theta) = \frac{\exp \sum_{v=0}^h Z_{jr}(\theta)}{\sum_{e=0}^{m_j} \exp \left[ \sum_{v=0}^e Z_{jr}(\theta) \right]}, k=0,1,2,\dots,m_j \dots\dots\dots(3)$$

dan

$$Z_{jh}(\theta) = Da_j(\theta - b_{jh}) = Da_j(\theta - b_j + d_h), b_{j0} = 0 \dots\dots\dots(4)$$

Dengan

- $P_{jk}(\theta)$  : probabilitas peserta berkemampuan  $\theta$  memperoleh skor kategori k pada butir j,
- $\theta$  : 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  $b_{jh}$  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  $\theta$  (van der Linden & Hambleton, 1997).

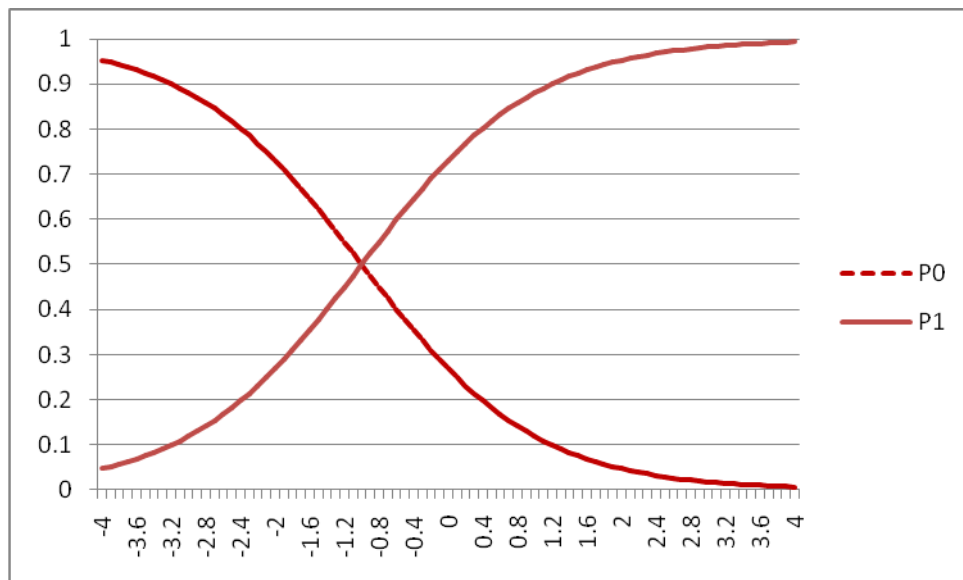
Jika  $\theta = b_{jk}$ , maka  $P_{jk}(\theta) = P_{jk-1}(\theta)$

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

Jika  $\theta < b_{jk}$ , maka  $P_{jk}(\theta) < P_{jk-1}(\theta)$ ,  $K=1,2,3,\dots,m_j \dots\dots\dots(5)$

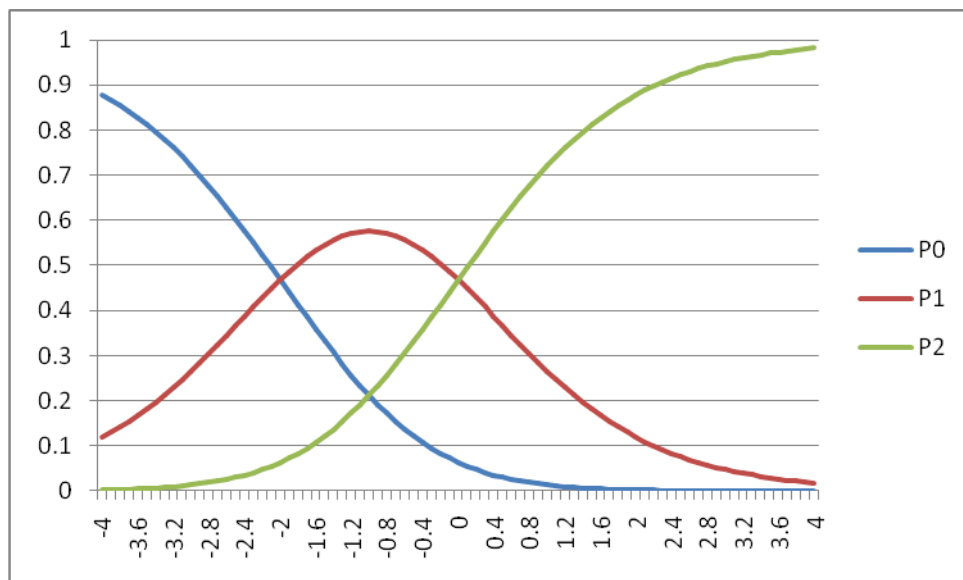
GPCM diformulasikan berdasarkan asumsi bahwa setiap probabilitas memilih kategori ke-k melampaui kategori ke-(k-1) dibangun oleh model dikotomi.  $P_{jk}$  merupakan probabilitas khusus memilih kategori ke-k dari  $m_j + 1$  kategori. Hubungan probabilitas menjawab benar untuk tiap kemampuan  $\theta$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

Pada dikotomus 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 1.



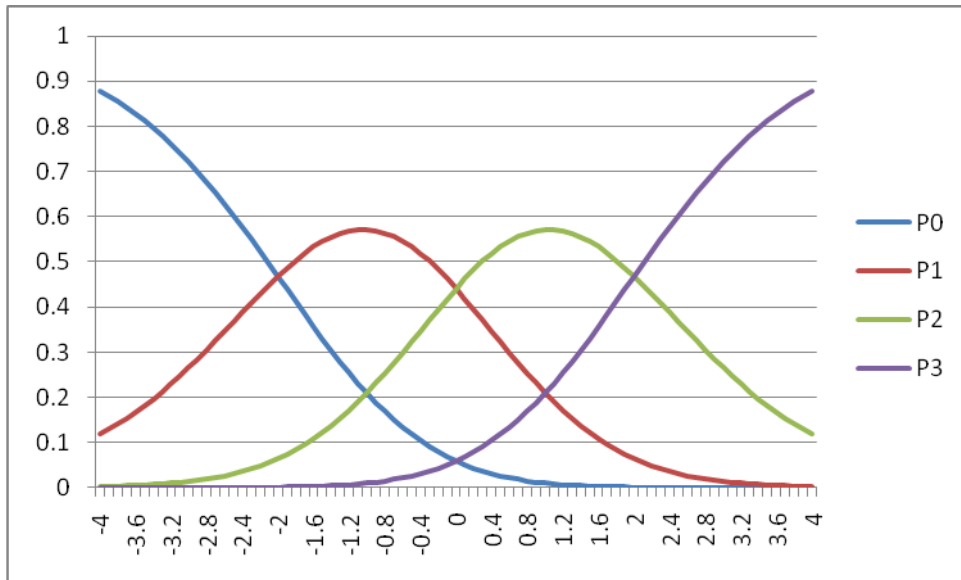
Gambar 1. Grafik CRF pada 2 kategori

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.



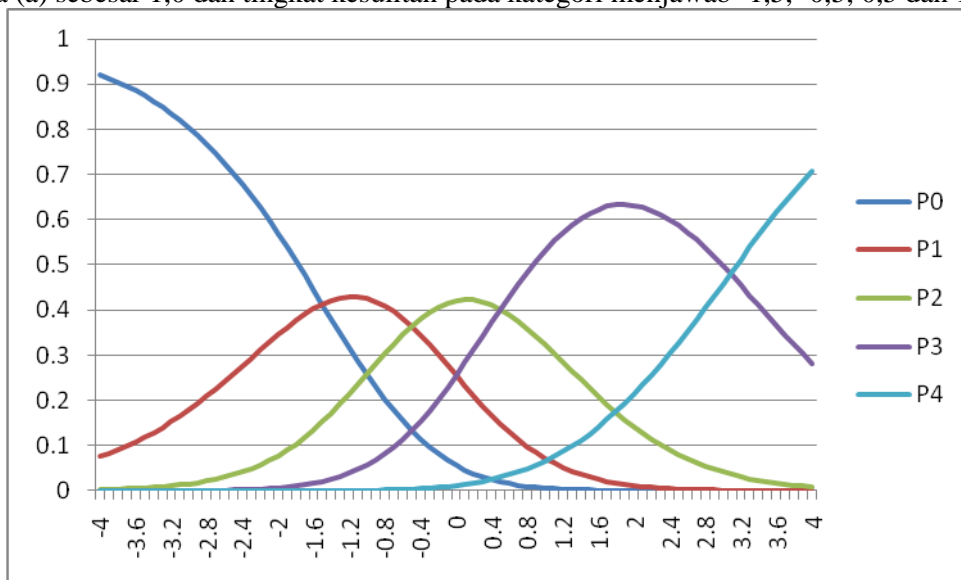
Gambar 2. Grafik CRF pada 3 kategori

Pada Gambar 3 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 3. Grafik CRF pada 4 kategori

Pada Gambar 4 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. Grafik CRF pada 5 Kategori

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  $\theta$  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  $\bar{r}_{ihf}$  dan ukuran sampel harapan provisional  $\bar{N}_f$  dihitung. Kemudian pada langkah maksimasi, diestimasi Marginal Maximum Likelihood (MML) dengan metode penskoran Fisher.

Program yang digunakan untuk mengestimasi parameter butir dan kemampuan diantaranya Parscale dari SSI (Muraki & Bock, 1997). Untuk dapat menggunakan program ini, ada 2 hal yang perlu menjadi perhatian yakni input data dan sintaks analisis. Input Data dapat menggunakan format teks \*.txt. Demikian pula sintaks analisis. Contoh masukan dan sintaks analisis disajikan pada Gambar 5 dan 6.

0001	42444232223343433332
0002	12221121122324121432
0003	32212212213342314121
0004	13222111113224221111
0005	21211122313132312131
0006	34443444344344444344
0007	23343213414312333331
0008	444444444444444344343
0009	44444234344444444333
0010	1111111111122111111
...	
0999	12112111112212111221
1000	23322322312244424442

Gambar 5. Contoh Masukan Analisis untuk GPCM

```

EXAMPL03.PSL - ARTIFICIAL EXAMPLE (MONTE CARLO DATA)
GENERALIZED PARTIAL CREDIT MODEL - EAP SCALE SCORES
>COMMENTS

>FILES  DFNAME='EXAMPL01.DAT', SAVE;
>SAVE   SCORE='EXAMPL03.SCO';
>INPUT  NIDW=4, NTOTAL=20, NTEST=1, LENGTH=20;
(4A1,10X,20A1)
>TEST   TNAME='SCALE1', ITEM=(1(1)20), NBLOCK=2;
>BLOCK1 BNAME='SBLOCK1', NITEMS=10, NCAT=4, SCORING=(1,2,3,4);
>BLOCK2 BNAME='SBLOCK2', NITEMS=10, NCAT=4, MODIFIED=(1,1,2,2),
SCOR=(1,2);
>CALIB  PARTIAL, LOGISTIC, NQPTS=15, CYCLE=(100,1,1,1,1),
NEWTON=2,
        CRIT=0.01;
>SCORE  MLE, SMEAN=0.0, SSD=1.0, NAME='PCR_MLE', PFQ=5;
    
```

Gambar 6. Contoh Sintaks Analisis untuk GPCM dengan Parscale dari SSI

Setelah melakukan *running* dari Parscale, diperoleh output yang berupa file \*.Ph1 (fase 1), \*.Ph2 (fase 2), dan \*.Ph3 (fase 3). Fase 1 merupakan hasil estimasi parameter berdasarkan teori tes klasik, yang meliputi proporsi menjawab benar perlangkah dan korelasi poliserial yang merupakan daya pembeda butir. Fase 2 merupakan fase estimasi parameter butir, dan fase 3 estimasi kemampuan peserta. Hasil tiap fase disajikan pada Gambar 7, 8, dan 9. Parameter butir

selengkapnya disajikan pada output \*.PAR, pada Gambar 10.

Ph1							
BLOCK NO.:		1	NAME: SBLOCK1				
ITEM	TOTAL	NOT PRESENT	OMIT	CATEGORIES			
				1	2	3	4
0001							
FREQ.	1000	0	0	194	303	313	190
PERC.		0.0	0.0	19.4	30.3	31.3	19.0
0002							
FREQ.	1000	0	0	204	284	310	202
PERC.		0.0	0.0	20.4	28.4	31.0	20.2
0003							
FREQ.	1000	0	0	206	308	285	201
PERC.		0.0	0.0	20.6	30.8	28.5	20.1
0004							
FREQ.	1000	0	0	199	301	286	214
PERC.		0.0	0.0	19.9	30.1	28.6	21.4
BLOCK ITEM	RESPONSE MEAN S.D.*	TOTAL SCORE MEAN S.D.*	PEARSON & POLYSERIAL CORRELATION	INITIAL SLOPE	INITIAL LOCATION		
SBLOCK1							
1 0001	2.499 1.009*	49.892 14.754*	0.778 0.830	1.488	-0.017		
2 0002	2.510 1.030*	49.892 14.754*	0.797 0.852	1.628	-0.036		
3 0003	2.481 1.031*	49.892 14.754*	0.785 0.839	1.545	0.013		
4 0004	2.515 1.037*	49.892 14.754*	0.805 0.861	1.695	-0.053		

Gambar 7. Output Phase 1

**Ph2**  
ITEM BLOCK 2 SBLOCK2

SCORING FUNCTION : 1.000 2.000  
STEP PARAMTER : 0.000 0.000  
S.E. : 0.000 0.000

ITEM	BLOCK	SLOPE	S.E.	LOCATION	S.E.	GUESSING	S.E.
0001	1	1.152	0.069	0.008	0.036	0.000	0.000
0002	1	1.192	0.076	-0.005	0.035	0.000	0.000
0003	1	1.157	0.076	0.030	0.036	0.000	0.000
0004	1	1.290	0.082	-0.011	0.034	0.000	0.000
0005	1	1.348	0.080	-0.006	0.034	0.000	0.000
0006	1	0.812	0.052	0.503	0.042	0.000	0.000
0007	1	0.886	0.058	0.515	0.039	0.000	0.000
0008	1	0.972	0.062	0.469	0.038	0.000	0.000
0009	1	0.876	0.056	0.493	0.039	0.000	0.000
0010	1	0.867	0.056	0.505	0.042	0.000	0.000
0011	2	0.899	0.072	-0.543	0.063	0.000	0.000
0012	2	0.882	0.069	-0.586	0.064	0.000	0.000
0013	2	1.088	0.085	-0.457	0.053	0.000	0.000
0014	2	0.964	0.076	-0.461	0.058	0.000	0.000
0015	2	0.989	0.081	-0.441	0.055	0.000	0.000
0016	2	0.766	0.061	-0.018	0.062	0.000	0.000
0017	2	0.614	0.055	-0.121	0.074	0.000	0.000
0018	2	0.755	0.062	0.077	0.063	0.000	0.000
0019	2	0.661	0.056	0.046	0.068	0.000	0.000
0020	2	0.687	0.059	0.070	0.066	0.000	0.000

Gambar 8. Output Phase 2

**Ph.3**

SUBJECT IDENTIFICATION WEIGHT/FREQUENCY

SCORE	NAME	GROUP	WEIGHT	MEAN	CATEGORY	ATTEMPTS	ABILITY	S.E.
0001			1		GROUP 01	1.00		
1	PCR_MLE	1	1.00	2.40	1.00	1.00	0.7728	0.1801
0002			2		GROUP 01	1.00		
1	PCR_MLE	1	1.00	1.45	1.00	1.00	-0.8738	0.1865
0003			3		GROUP 01	1.00		
1	PCR_MLE	1	1.00	1.65	1.00	1.00	-0.4798	0.1689
0004			4		GROUP 01	1.00		
1	PCR_MLE	1	1.00	1.35	1.00	1.00	-0.9355	0.1907
0005			5		GROUP 01	1.00		
1	PCR_MLE	1	1.00	1.50	1.00	1.00	-0.7876	0.1814

Gambar 9. Output Phase 3



```
.par
EXAMPL03.PSL - ARTIFICIAL EXAMPLE (MONTE CARLO DATA)
      GENERALIZED PARTIAL CREDIT MODEL - EAP SCALE SCORES
TEST 01  2  20  7  0  1
      10 10
GROUP 01
SBLOCK1  40001  1.15154  0.06889  0.00831  0.03576  0.00000  0.00000
SBLOCK1  40002  1.19216  0.07569  -0.00498  0.03499  0.00000  0.00000
SBLOCK1  40003  1.15734  0.07614  0.03038  0.03561  0.00000  0.00000
.....
SBLOCK2  20011  0.89922  0.07164  -0.54271  0.06285  0.00000  0.00000
SBLOCK2  20012  0.88179  0.06943  -0.58600  0.06411  0.00000  0.00000
SBLOCK2  20013  1.08803  0.08457  -0.45680  0.05306  0.00000  0.00000
```

Gambar 10. Output Estimasi Parameter (\*.PAR)

Hasil estimasi kemampuan disajikan pada skala interval (-4,+4). Agar dapat dimanfaatkan untuk memperoleh interpretasi yang lebih baik, selanjutnya perlu dilakukan dengan transformasi linear biasa. Misalnya disajikan pada rentang [0,10] atau [0,100].

## KESIMPULAN

Pada tes esai atau constructed response, penskoran dilakukan dengan politomus. Untuk mengestimasi kemampuan, model alternatif yang dapat dilakukan yakni dengan *Generalized Partial Credit Model* (GPCM). Estimasi kemampuan dapat dilakukan dengan bantuan Parscale dari SSI. Estimasi kemampuan disajikan pada fase 3 pada skala (-4,+4) yang kemudian dapat ditransformasi agar lebih mudah dimaknai. Untuk dapat mengetahui ketepatannya dalam mengestimasi kemampuan, perlu dilakukan studi simulasi antara yang menggunakan total skor menjawab benar dan yang menggunakan GPCM.

## DAFTAR PUSTAKA

- Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: SSI.
- Gronlund, N.E. & Linn, R.L. (1990). *Measurement and evaluation in teaching (6th ed)*. New York : Collier Macmillan Publishers
- 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.
- 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.
- 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.

*Heri Retnawati / Mengestimasi Kemampuan Peserta*

Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.