

Penerapan *Partial Credit Model* (PCM) dalam Mengevaluasi Tes Uraian

Application of Partial Credit Model (PCM) in Evaluating the Essay Tes

Ni Kadek Rini Purwati^{a,*}, I Wayan Sumandya^b

Prodi. Pendidikan Matematika FPMIPA IKIP PGRI Bali

*Pos-el: rini@ictp.ac.id

Abstrak. Penelitian dengan judul “Penerapan *Partial Credit Model* (PCM) dalam Mengevaluasi Tes Uraian” ini bertujuan untuk mengetahui bagaimana *Partial Credit Model* (PCM) diterapkan dalam mengevaluasi tes. Penelitian ini merupakan penelitian deskriptif eksploratif dan analisis data dilakukan menggunakan PCM dengan bantuan software Parscale. Data penelitian diperoleh dari hasil ujian tengah semester mata kuliah metode numerik.

Penerapan PCM dalam mengevaluasi tes uraian memberikan informasi bahwa variabel kemampuan peserta tes berpengaruh signifikan terhadap peluang menjawab benar, yakni sebesar 98,1%, rata-rata tingkat kesukaran adalah sebesar -0,078 yang mengindikasikan bahwa tes metode numerik termasuk dalam kategori sukar, rata-rata kemampuan peserta tes adalah -0,008 yang menunjukkan bahwa rata-rata kemampuan peserta dalam menjawab tes metode numerik berada dalam kategori sedang, dan berdasarkan fungsi informasi dan *Standard Error of Measurement* (SEM) diperoleh bahwa tes metode numerik yang dikembangkan cocok untuk mahasiswa dengan kemampuan dalam kategori sedang.

Kata-Kata Kunci: *Partial Credit Model* (PCM), tingkat kesukaran, tes uraian.

Abstract. The study, titled "Application of Partial Credit Model (PCM) in Evaluating the Essay Test" aims to find out how the Partial Credit Model (PCM) is applied in evaluating tests. This research is an explorative descriptive research and data analysis is done using PCM with the help of Parscale software. Research data obtained from the results of the middle tests of numerical method.

The application of PCM in evaluating the essay test provided information that the variable ability of test participants significantly influenced to the probability of correct answer, ie 98.1%, the average of the difficulty level was -0.078 indicating that the numerical method test included in the difficult category, the average ability of the test participants was -0.008 indicating that the average ability of the participants in answering the numerical methods test was in the medium category, and based on the information function and the Standard Error of Measurement (SEM) it was found that the numerical method test developed was suitable for students with the ability to medium category.

Key Words: Partial Credit Model (PCM), difficulty level, essay test

PENDAHULUAN

Tes adalah suatu instrumen atau prosedur yang sistematis untuk mengukur suatu perilaku tertentu (Gronlund, 1985). Berdasarkan definisi tersebut, maka tes dapat digunakan

untuk mengestimasi kemampuan peserta didik. Suatu tes haruslah dapat mengukur hasil belajar yang sesuai dengan tujuan instruksional serta mempunyai reliabilitas dan validitas

yang baik sehingga hasilnya dapat ditafsirkan dengan tepat.

Tes dapat dibedakan menjadi dua macam, yakni tes subjektif yang pada umumnya berbentuk esay (uraian) dan tes objektif yang dalam pemeriksaannya dapat dilakukan secara objektif. Tes yang baik akan dapat mengestimasi dengan baik kemampuan peserta didik. Salah satu tes yang dapat memberikan kesempatan kepada peserta didik untuk mengekspresikan jawaban, ide-ide, serta mengorganisasi jawaban sendiri adalah tes uraian. Melalui tes uraian, peserta didik dituntut dapat menganalisis permasalahan yang diberikan dan mengaitkannya dengan konsep-konsep yang ada.

Tes yang disusun merupakan dasar utama untuk memberikan skor kepada peserta didik. Skor adalah ukuran tentang prestasi belajar peserta didik sesuai dengan tujuan instruksional. Hasil penskoran ini dapat digunakan sebagai dasar melakukan analisis item. Analisis item (analisis butir soal) dapat memberikan informasi tentang kualitas sebuah item dan petunjuk untuk mengadakan perbaikan. Analisis item dilakukan dengan menganalisis respon/tanggapan siswa terhadap item tersebut.

Berkaitan dengan penskoran, maka tes uraian perlu dibarengi dengan teknik penskoran yang tepat mengingat tes ini merupakan suatu tes subjektif yang proses pemberian skornya dipengaruhi oleh opini atau penilaian dari pendidik atau pemeriksa tes tersebut. Umumnya, penskoran tes uraian dilakukan berdasarkan langkah-langkah yang ditempuh untuk menjawab suatu butir soal. Penskoran dilakukan perlangkah dan skor perbutir diperoleh dengan menjumlah skor tiap langkah. Akan tetapi teknik penskoran ini tidak

memperhatikan tingkat kesukaran yang berbeda dalam tiap langkah, sehingga hasil yang diperoleh kurang akurat digunakan dalam mengestimasi kemampuan peserta tes. Untuk itu diperlukan suatu teknik penskoran yang lebih tepat yang dapat mengevaluasi karakteristik butir soal meliputi daya beda dan tingkat kesukaran, dan sekaligus dapat mengestimasi kemampuan peserta tes melalui respon terhadap butir soal. Salah satu teori yang dapat digunakan adalah teori respon butir untuk penskoran politomus yang salah satunya dengan *Partial Credit Model* (PCM).

Partial Credit Model (PCM) adalah teori respon butir politomus yang berguna dalam mengevaluasi suatu tes baik karakteristik tes maupun kemampuan peserta tes. Parameter butir yang diestimasi pada PCM adalah tingkat kesukaran butir tes. Tingkat kesukaran butir ini merupakan salah satu karakteristik tes. PCM yang merupakan model 1-PL (1 parameter logistik) mengasumsikan daya beda setiap butir adalah sama dan tingkat kesukaran dalam setiap tahap tidak perlu terurut. Hal ini seperti dinyatakan dalam Masters (dalam Wetzel, 2014):

“In the partial credit model . . . the item parameters $\delta_{i1}, \delta_{i2}, \dots, \delta_{im}$ govern the transitions between adjacent response categories. Order is not incorporated through the values of these locally defined parameters, which are in fact free to take any values at all.”

Selain informasi mengenai karakteristik tes, hasil analisis PCM dapat memberikan beberapa informasi terkait sebaran kemampuan peserta tes, hubungan antara kemampuan peserta tes dengan peluang menjawab benar, serta fungsi informasi dan *Standard Error of Measurement* (SEM). Penggunaan PCM

dalam mengevaluasi tes diharapkan dapat memberikan informasi yang berguna dalam mengembangkan suatu tes menjadi instrumen yang dapat mengukur kemampuan peserta tes secara lebih akurat. Berdasarkan penjabaran tersebut, maka dirasa perlu untuk melakukan penelitian terkait “Penerapan *Partial Credit Model* (PCM) dalam Mengevaluasi Tes Uraian”.

METODE PENELITIAN

Penelitian ini termasuk jenis penelitian deskriptif eksploratif dengan subjek penelitiannya adalah mahasiswa semester VI Jurusan Pendidikan Matematika IKIP PGRI Bali. Objek yang dipilih adalah soal ujian tengah semester mata kuliah metode numerik yang berupa tes uraian.

Data dari tes uraian dianalisis menggunakan *Partial Credit Model* (PCM) dengan bantuan software *Parscale* dari *SSi*. Langkah-langkah yang dilakukan dalam analisis data, yakni:

- 1) Mengumpulkan data respon butir soal ujian tengah semester mata kuliah metode numerik
- 2) Skoring data
- 3) Menyusun sintaks analisis pada *Parscale*
- 4) *Running Parscale* untuk menghasilkan output
- 5) Mengestimasi indeks parameter butir, yaitu tingkat kesukaran butir dan tingkat kesukaran tahap tiap butir berdasarkan output *Parscale*
- 6) Mengestimasi tingkat kemampuan dan peluang mahasiswa menjawab butir soal dengan benar

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengetahui penerapan *Partial Credit*

Model (PCM) dalam mengevaluasi tes uraian. Evaluasi tes menggunakan PCM dilakukan dengan bantuan software *Parscale* dari *SSi*. Evaluasi tes dilakukan setelah hasil jawaban peserta tes metode numerik diberikan skor. Berikut ini disajikan contoh proses skoring respon butir tes metode numerik pada butir 1 peserta 1, peserta 4, dan peserta 5.

Tabel 1.
Contoh Proses Skoring

Peserta $a(n)$	<i>Ordered Performance Level</i> (Kategori)					Skor
	0	1	2	3	4	
	y_{0ni}	y_{1ni}	y_{2ni}	y_{3ni}	y_{4ni}	x_{ni}
1	1	1	1	1	1	4
4	1	1	1	1		3
5	1					0

Berdasarkan Tabel 1, simbol $y_{0ni}, y_{1ni}, y_{2ni}, y_{3ni}, y_{4ni}$ masing-masing menunjukkan langkah penyelesaian ke-0, ke-1, ke-2, ke-3, dan ke-4 pada butir i yang diselesaikan siswa n dan x_{ni} merupakan skor kategori/banyaknya langkah penyelesaian yang telah diselesaikan oleh peserta n pada butir i . Pada tabel terlihat bahwa peserta 1 mendapat nilai 1 pada setiap langkah penyelesaian. Hal ini menunjukkan bahwa peserta 1 dapat menjawab dengan benar setiap langkah penyelesaian, sedangkan peserta 4 hanya dapat menyelesaikan dengan benar hingga langkah ke 3 dan peserta 5 tidak dapat menyelesaikan dengan benar setiap langkah.

Data hasil skoring yang telah diperoleh kemudian dievaluasi menggunakan PCM dengan sintaks sebagai berikut:

```

PARTIAL CREDIT MODEL
  Numerical test, ONE SAMPLE (N=79,
N=(0,1)), 4 ITEMS
>COMMENTS
  This data set was first numerical
method test
>FILES  DFNAME='math.DAT', SAVE;
>SAVE   SCORE='math.EAP';
>INPUT  NIDW=4, NTOT=4, NTEST=1,
LENGTH=(4);
(4A1,1X,4A1)
>TEST   ITEM=(1(1)4), NBLOCK=1;
>BLOCK1
NITEMS=4, NCAT=5, ORIGINAL=(0,1,2,3,4);
>CALIB  LOGISTIC, PARTIAL, NQPTS=21,
CYCLES=(100,1,1,1,1), NEWTON=2,
CRIT=0.005, SCALE=1.7;
>SCORE  EAP, NAME=TEST1;

```

Gambar 1. Sintaks Analisis PCM

Setelah sintaks analisis tersusun, selanjutnya adalah *running* program untuk menghasilkan output. *Running* akan terjadi bertahap dimulai dari *Phase* nol hingga *Phase* tiga.

Output *Phase* 0 menampilkan informasi mengenai banyaknya butir tes dan metode estimasi, seperti disajikan pada Gambar 2.

```

*** POLYTOMOUS MODEL ITEM ANALYSER ***
*** PHASE 0 ***
MODEL SPECIFICATIONS
LOGISTIC - PARTIAL CREDIT ITEM RESPONSE MODEL IS SPECIFIED.
SCALE CONSTANT 1.70 FOR SLOPE PARAMETERS.
CALIBRATION PARAMETERS
MAXIMUM NUMBER OF EM CYCLES:          100
PARAMETERS FOR SCORING AND TEST AND ITEM INFORMATION
METHOD OF SCORING SUBJECTS:           EXPECTATION A
POSTERIORI                             (EAP; BAYES)
ESTIMATES)
NUMBER OF ITEMS :      4
INPUT AND RECODED RESPONSE OF FIRST AND SECOND OBSERVATIONS
OBSERVATION #      1
GROUP: 1
ID: 0001
INPUT RESPONSES:   4 4 3 1
RECODED RESPONSES: 5 5 4 2
OBSERVATION #      2
GROUP: 1
ID: 0002
INPUT RESPONSES:   4 4 3 0
RECODED RESPONSES: 5 5 4 1
NORMAL END

```

Gambar 2. Output Phase 0

Metode estimasi (*method of scoring subject*) yang digunakan adalah *Expected A Posteriori (EAP) Bayes* dan banyak butir tes yang diestimasi adalah 4.

Output *Phase* 1 menampilkan informasi mengenai hasil estimasi parameter berdasarkan teori tes klasik, yang meliputi proporsi peserta tes terhadap respon butir perkategori, mean dan standar deviasi untuk setiap butir/item, serta korelasi *pearson* dan *polyserial*.

```

*** PHASE 1 ***
SUMMARY ITEM STATISTICS
BLOCK NO.: 1 NAME: BLOCK
ITEM | TOTAL | NOT PRESENT | OMIT | CATEGORIES
      |       |             |      | 1     2     3     4     5
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
0001 | 79    | 0          | 0    | 55    0    1    2    21
FREQ. |       | 0.0        | 0.0  | 69.6  0.0  1.3  2.5  26.6
PERC. |       |            |      |
-----|-----|-----|-----|-----|-----|-----|-----|
0002 | 79    | 0          | 0    | 14    7    16   6    36
FREQ. |       | 0.0        | 0.0  | 17.7  8.9  20.3  7.6  45.6
PERC. |       |            |      |
-----|-----|-----|-----|-----|-----|-----|
0003 | 79    | 0          | 0    | 22    5    5    39   8
FREQ. |       | 0.0        | 0.0  | 27.8  6.3  6.3  49.4  10.1
PERC. |       |            |      |
-----|-----|-----|-----|-----|-----|
0004 | 79    | 0          | 0    | 26    2    3    3    45
FREQ. |       | 0.0        | 0.0  | 32.9  2.5  3.8  3.8  57.0
PERC. |       |            |      |
-----|-----|-----|-----|-----|-----|
CUMMUL. |       |       |       | 117   14   25   50   110
FREQ. |       |       |       | 37.0  4.4  7.9  15.8  34.8
PERC. |       |       |       |
-----|-----|-----|-----|-----|-----|
BLOCK ITEM | RESPONSE MEAN | TOTAL SCORE MEAN | PEARSON & POLYSERIAL CORRELATION | INITIAL SLOPE | INITIAL LOCATION
-----|-----|-----|-----|-----|-----|
BLOCK 1 0001 | 2.165 | 12.278 | 0.731 | 1.000 | -1.761
      | 1.782* | 5.059* | 0.950 |       |
      | 3.544 | 12.278 | 0.751 | 1.560 | -0.322
BLOCK 2 0002 | 1.549* | 5.059* | 0.842 |       |
      | 3.076 | 12.278 | 0.792 | 2.050 | 0.340
      | 1.439* | 5.059* | 0.899 |       |
BLOCK 4 0004 | 3.494 | 12.278 | 0.787 | 1.000 | -0.074
      | 1.848* | 5.059* | 0.962 |       |
-----|-----|-----|-----|-----|-----|
CATEGORY | SCORING | MEAN | S.D. | PARAMETER
-----|-----|-----|-----|-----|
1 | 1.000 | 7.349 | 3.778 | -0.144
2 | 2.000 | 10.071 | 4.574 | -0.144
3 | 3.000 | 11.440 | 3.371 | -0.508
4 | 4.000 | 15.060 | 3.082 | -0.014
5 | 5.000 | 16.091 | 3.164 | -0.350
NORMAL END

```

Gambar 3. Output Phase 1

Berdasarkan output phase 1 pada Gambar 3 diperoleh nilai korelasi *pearson* untuk setiap butir lebih besar dari 0,2 ($r_{hitung} > r_{tabel}$) sehingga dapat disimpulkan bahwa butir tes valid. *Initial location* menunjukkan nilai rata-rata parameter ambang batas (*threshold*) dari setiap butir tes yang dalam hal ini parameter ambang batas (*threshold*) adalah tingkat kesukaran (*b*). Berdasarkan data terlihat bahwa *initial location* butir 1 adalah -1,761, *initial location* butir 2 adalah -0,322, *initial location* butir 3 adalah 0,340, *initial location* butir 4 adalah -0,074.

Output *Phase 2* berisi rincian tentang estimasi parameter butir tes, seperti disajikan pada Gambar 4.

```

*** PHASE 2 ***
[E-M CYCLES] PARTIAL CREDIT MODEL
STEP AND ITEM PARAMETERS AFTER CYCLE 0
LARGEST CHANGE= 0.000
-2 LOG LIKELIHOOD = 1889.335
STEP AND ITEM PARAMETERS AFTER CYCLE 1
LARGEST CHANGE= 0.694 ( 1.000-> 0.306) at Slope of
Item: 4 0004
-2 LOG LIKELIHOOD = 1356.644

ITEM BLOCK 1 BLOCK
SCORING FUNCTION : 1.000 2.000 3.000
4.000 5.000
STEP PARAMETER : 0.000 -2.175 0.999
0.399 0.777
S.E. : 0.000 0.463 0.524
0.348 0.227

+-----+
| ITEM | BLOCK | SLOPE | S.E. | LOCATION | S.E. | GUESSING |
| S.E. |
+-----+
| 0001 | 1 | 1.171 | 0.280 | 0.575 | 0.170 | 0.000 |
| 0.000 |
| 0002 | 1 | 0.293 | 0.053 | -0.537 | 0.235 | 0.000 |
| 0.000 |
| 0003 | 1 | 0.247 | 0.061 | -0.028 | 0.294 | 0.000 |
| 0.000 |
| 0004 | 1 | 0.637 | 0.095 | -0.323 | 0.165 | 0.000 |
| 0.000 |
+-----+

SUMMARY STATISTICS OF PARAMETER ESTIMATES
+-----+
| PARAMETER | MEAN | STN DEV | N |
+-----+
| SLOPE | 0.587 | 0.426 | 4 |
| LOG(SLOPE) | -0.729 | 0.721 | 4 |
| THRESHOLD | -0.078 | 0.483 | 4 |
| GUESSING | 0.000 | 0.000 | 0 |
+-----+

NORMAL END

```

Gambar 4. Output Phase 2

Berdasarkan Gambar 4 diperoleh informasi bahwa butir 1 memiliki tingkat kesukaran sebesar 0,575, butir 2 memiliki tingkat kesukaran sebesar -0,537, butir 3 memiliki tingkat kesukaran sebesar -0,028, dan butir 4 memiliki tingkat kesukaran sebesar -0,323. Rata-rata tingkat kesukaran (*threshold*) sebesar -0,078 yang mengindikasikan bahwa tes termasuk dalam kategori sukar.

Output *Phase 3* menyajikan data mengenai informasi skoring tes dan estimasi kemampuan peserta tes, seperti disajikan pada Gambar 5.

```

[ PHASE 3 ]
SCORING OF MAIN TEST
TEST 01

MEANS AND STANDARD DEVIATIONS OF ABILITY DISTRIBUTIONS

SCORE      MEAN      STANDARD      TOTAL
NAME              DEVIATION      FREQUENCIES
-----
TEST1      -0.008    0.875        79.00
-----

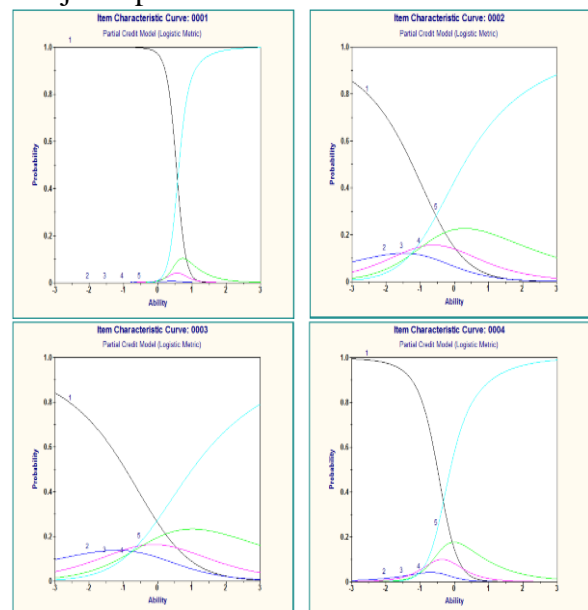
NORMAL END

```

Gambar 5. Output Phase 3

Berdasarkan Gambar 5 diperoleh rata-rata kemampuan peserta tes adalah -0,008 yang menunjukkan bahwa rata-rata kemampuan peserta dalam menjawab tes metode numerik berada dalam kategori sedang.

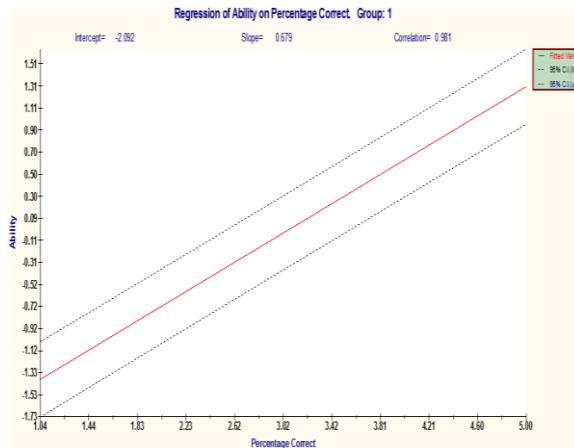
Selain output dari tiap phase, *Parscale* juga menghasilkan output berupa Kurva Karakteristik Item (*Item Characteristic Curve/ICC*), seperti disajikan pada Gambar 6.



Gambar 6. Kurva ICC

Titik potong pada kurva ICC menunjukkan tingkat kesukaran tahap yang berkaitan dengan transisi dari satu kategori ke kategori berikutnya.

Hubungan antara kemampuan peserta tes dengan peluang menjawab benar disajikan pada Gambar 7.



Gambar 7. Hubungan antara Kemampuan Peserta Tes dengan Peluang Menjawab Benar

Hasil evaluasi PCM pada Gambar 6 menunjukkan bahwa variabel kemampuan peserta tes berpengaruh signifikan terhadap peluang menjawab benar tes, yakni sebesar 98,1% ($Correlation=0,981$). Hal ini menandakan bahwa jika kemampuan peserta tes tinggi, maka peluang menjawab benar juga tinggi dan begitupun sebaliknya.

Berdasarkan fungsi informasi dan *Standard Error of Measurement* (SEM) yang juga dihasilkan dari analisis PCM diperoleh bahwa tes metode numerik yang dikembangkan cocok untuk mahasiswa dengan kemampuan dalam kategori sedang, yakni kemampuan yang berkisar pada rentang -0,8 sampai 1.

PCM dapat digunakan untuk mengevaluasi tes uraian dan sekaligus menyediakan teknik penskoran tes uraian yang lebih tepat. Evaluasi tes uraian ini memberikan informasi mengenai karakteristik butir soal yang dalam hal ini adalah tingkat kesukaran butir, hubungan antara peluang menjawab benar dengan kemampuan peserta tes, serta estimasi kemampuan peserta tes.

PCM adalah teori respon butir politomus yang pada awalnya dikembangkan untuk menganalisis butir tes yang memerlukan beberapa langkah penyelesaian dimana kredit parsial dapat diberikan pada langkah-langkah yang dapat diatasi oleh individu (Widhiarso, 2010). Hal ini seperti dinyatakan juga oleh Masters (dalam Wetzel, 2014) bahwa:

“The PCM is a polytomous item response model that assumes ordered response categories as they exist in partial credit items (incorrect, partially correct, fully correct) or in questionnaires using unidimensional rating scales (e.g., strongly disagree to strongly agree.”

Sistem penskoran kredit parsial biasa digunakan dalam ujian tertulis yang memungkinkan penguji untuk mengidentifikasi pemahaman secara parsial.

Menurut Adam (dalam Wetzel, 2014), PCM tidak menempatkan suatu hubungan antara tingkat kesukaran yang terurut dengan kategori respon yang terurut. Tingkat kesukaran butir saling terkait dengan kemampuan peserta tes dimana semakin tinggi tingkat kesukaran butir, maka semakin tinggi kemampuan peserta tes dalam menjawab butir secara benar. Hal ini menunjukkan adanya hubungan antara peluang menjawab benar dengan kemampuan peserta tes. Berbeda halnya dalam teori tes klasik dimana karakteristik butir bergantung pada level kemampuan peserta tes. Jika butir diselesaikan oleh peserta dengan kemampuan tinggi maka butir menunjukkan tingkat kesukaran rendah, sebaliknya untuk peserta dengan kemampuan rendah dan sedang maka butir menunjukkan tingkat kesukaran tinggi.

Meskipun PCM memerlukan langkah penyelesaian yang terurut tetapi tingkat kesukaran tiap langkah tidak perlu terurut. Hal ini seperti dinyatakan oleh Widhiarso (2010) bahwa tingkat kesukaran butir untuk satu tahap independen dengan tahap-tahap lainnya. Informasi mengenai tingkat kesukaran tahap dapat dilihat pada titik potong kurva ICC, yakni tingkat kesukaran tidak berurutan pada setiap tahap dimana suatu tahap dapat lebih sulit dibandingkan tahap berikutnya.

Berdasarkan hasil analisis data diperoleh informasi terkait tingkat kesukaran perbutir, yakni butir 1 memiliki tingkat kesukaran sedang (0,575), butir 2 memiliki tingkat kesukaran tinggi (-0,537), butir 3 memiliki tingkat kesukaran tinggi (-0,028), dan butir 4 memiliki tingkat kesukaran tinggi (-0,323). Hal ini sesuai dengan kategori tingkat kesukaran menurut Aziz (2015), yakni: a) tingkat kesukaran tinggi jika nilai parameter tingkat kesukaran kurang dari 0,30; b) tingkat kesukaran sedang jika nilai parameter tingkat kesukaran berkisar pada rentang nilai 0,30 sampai 0,70; c) tingkat kesukaran rendah jika nilai parameter tingkat kesukaran lebih dari 0,70. Rata-rata tingkat kesukaran (*threshold*) pada tes metode numerik ini adalah sebesar -0,078 dengan simpangan baku sebesar 0,483 yang mengindikasikan bahwa tes metode numerik termasuk dalam kategori sukar. Hasil estimasi tingkat kesukaran ini juga menunjukkan bahwa butir tes termasuk dalam kriteria baik untuk digunakan karena nilai parameter tingkat kesukaran berada pada rentang $-2 < b < 2$ (Hambleton & Swaminathan, 1985).

Estimasi tingkat kesukaran tahap dapat digunakan untuk melihat estimasi

kemampuan peserta tes. Seperti dinyatakan dalam Aziz (2015), tingkat kesukaran adalah proporsi peserta tes yang menjawab benar. Hasil estimasi kemampuan peserta tes berdasarkan analisis PCM, yakni rata-rata kemampuan peserta tes adalah -0,008 yang menunjukkan bahwa rata-rata kemampuan peserta dalam menjawab tes metode numerik berada dalam kategori sedang. Rentang kemampuan peserta tes berkisar pada $-3 \leq \theta \leq 3$ dan dapat dikelompokkan dalam 3 kategori, yakni rentang $-3 \leq \theta < -1$ termasuk kategori kemampuan rendah, rentang $-1 \leq \theta \leq 1$ termasuk kategori kemampuan sedang, dan rentang $1 < \theta \leq 3$ termasuk kategori kemampuan tinggi.

PCM mengevaluasi suatu tes berdasarkan hasil estimasi parameter butir dan estimasi kemampuan peserta tes. Estimasi kemampuan dan estimasi parameter butir dapat dilakukan bersamaan, yakni langkah pertama adalah melakukan estimasi parameter butir dan hasilnya digunakan untuk estimasi parameter kemampuan. PCM memprediksi kemampuan peserta melalui kemampuan mereka dalam menjawab butir tes secara benar dimana semakin tinggi kemampuan peserta, maka semakin tinggi peluang menjawab benar.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan analisis *Partial Credit Model* (PCM) yang telah dilakukan diperoleh kesimpulan bahwa evaluasi tes uraian menggunakan PCM dilakukan setelah hasil jawaban peserta tes diberikan skor kategori. Penerapan PCM dalam mengevaluasi tes uraian memberikan informasi mengenai tingkat kesukaran butir, hubungan antara peluang

menjawab benar dengan kemampuan peserta tes, serta estimasi kemampuan peserta tes. Berikut ini informasi yang diperoleh:

- a. Rata-rata tingkat kesukaran (*threshold*) adalah sebesar $-0,078$ dengan simpangan baku sebesar $0,483$ yang mengindikasikan bahwa tes metode numerik termasuk dalam kategori sukar.
- b. Variabel kemampuan peserta tes berpengaruh signifikan terhadap peluang menjawab benar, yakni sebesar $98,1\%$ ($Correlation=0,981$) yang menandakan bahwa jika kemampuan peserta tes tinggi, maka peluang menjawab benar juga tinggi dan begitupun sebaliknya.
- c. Rata-rata kemampuan peserta tes adalah $-0,008$ yang menunjukkan bahwa rata-rata kemampuan peserta dalam menjawab tes metode numerik berada dalam kategori sedang.
- d. Berdasarkan fungsi informasi dan *Standard Error of Measurement* (SEM) diperoleh bahwa tes metode numerik yang dikembangkan cocok untuk mahasiswa dengan kemampuan dalam kategori sedang. Berdasarkan analisis dan kesimpulan yang diperoleh, maka disarankan agar informasi mengenai hasil evaluasi tes uraian dengan menggunakan PCM dapat dimanfaatkan sebagai umpan balik dalam merevisi dan mengembangkan instrumen tes yang ada. Selain itu, teknik skoring dengan kredit parsial dapat diterapkan pada jenis tes uraian lainnya.

DAFTAR RUJUKAN

Aziz, Abdul. (2015). Karakteristik Soal Ujian Akhir Semester Gasal Mata Pelajaran Fisika SMA Kelas X di Kabupaten Lombok Tengah Nusa

Tenggara Barat. *Prosiding Seminar Nasional Fisika (E-Journal) SNF2015 Volume IV*.

Du Toit, M. (Ed.). (2003). *IRT from SSI : BILOG-MG, MULTILOG, PARSCALE, TES TFACT*.

Chicago: Scientific Software International.

Gronlund, Norman. E. (1985). *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company.

Hambleton & Swaminathan.(1985). *Item Response Theory Principles and Applications*. Boston: Kluwer Nijhoff Publishing.

Istiyono, Edi., Djemari Mardapi, Suparno. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan Volume 18(1)*.

Kurniawan, Dwi dan Devi. (2015). Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*.

Masters, Geoff N. 1982. A Rasch Model For Partial Credit Scoring. *Psychometrika 47(2)*.

Penfield, Randall David. (2014). An NCME Instructional Module on Polytomous Item Response Theory Models. *Educational Measurement: Issues and Practice 33(1):36-48*.

Retnawati, Heri. (2014). *Teori Respons Butir dan Penerapannya untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika.

Retnawati, Heri. (2015). Karakteristik Butir Tes dan Analisisnya. *Disampaikan pada*

- Workshop Penyusunan Instrumen Uji Kompetensi Guru, 30 September-1 Oktober 2015 Di PPP4TK Seni Budaya Yogyakarta.*
- Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan dan Kebudayaan* 17(6).
- Wahyu, Widhiarso. (2010). *Model Politomi dalam Teori Respons Butir*. Fakultas Psikologi UGM.
- Wasis. (2011). Model Penskoran Partial Credit pada Butir Multiple True-False Bidang Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan* 15(1).
- Wetzel, Eunike., Claus H. Carstensen. (2014). Reversed Thresholds in Partial Credit Models: A Reason for Collapsing Categories. *Assessment* 21(6), 765-774.