

## **ANALISIS DATA UJIAN NASIONAL MATEMATIKA BERDASARKAN PENSKORAN MODEL RASCH DAN MODEL PARTIAL CREDIT**

**Awal Isgiyanto**

*Jurusan Matematika Universitas Bengkulu*

### **Abstrak**

Penelitian ini bertujuan untuk menemukan: (1) atribut-atribut yang mendasari butir-butir soal Ujian Nasional Matematika, (2) ketuntasan atribut yang mendasari butir-butir soal Ujian Nasional Matematika; (3) kemampuan peserta dalam matematika pada model *Rasch* dan model *Partial Credit*; dan (4) nilai informasi butir dan informasi tes pada Ujian Nasional Matematika. Penelitian ini merupakan penelitian deskriptif eksploratif, yang menerapkan diagnosis *post-hoc*. Diagnosis ini digambarkan sebagai pendekatan *retrofitting* melalui analisis butir soal dan respons butir pada Ujian Nasional Matematika untuk menemukan informasi pada kategori isi, proses, dan keterampilan. Subjek penelitian adalah seluruh peserta Ujian Nasional Matematika SMP di wilayah Kabupaten Bantul, Daerah Istimewa Yogyakarta Tahun Pelajaran 2007/2008. Objek yang dipilih adalah butir soal, dan respons butir peserta Ujian Nasional Matematika yang diskor model *Rasch* dan model *Partial Credit*. Hasil penelitian menunjukkan bahwa: (1) atribut-atribut yang mendasari butir soal matematika ada 47 atribut, meliputi 4 atribut isi, 36 atribut proses, dan 7 atribut keterampilan, (2) ketuntasan atribut kategori isi, proses, dan keterampilan pada penskoran model *Partial Credit* lebih tinggi dan lebih akurat dibandingkan dengan ketuntasan atribut pada penskoran model *Rasch*, (3) rerata kemampuan peserta dalam matematika pada model *Partial Credit* lebih tinggi dibandingkan dengan rerata kemampuan peserta pada model *Rasch*, dan (4) informasi butir dan informasi tes pada penskoran model *Partial Credit* lebih baik dibandingkan dengan informasi butir dan informasi tes pada model *Rasch*.

**Kata kunci:** matematika, *retrofitting*, model *Rasch*, model *Partial Credit*

## **PENDAHULUAN**

### **Latar Belakang**

Pengukuran hasil belajar yang dilakukan oleh pemerintah diwujudkan dalam bentuk Ujian Nasional. Ujian Nasional (UN) adalah kegiatan pengukuran pencapaian kompetensi peserta didik pada beberapa mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi dalam rangka menilai pencapaian Standar Nasional Pendidikan. (Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20 Tahun 2007, 2007a). Penyelenggara UN mempunyai harapan agar semua pihak yang terkait dengan UN terdorong bekerja keras untuk mencapai hasil UN yang sebaik-baiknya.

Ujian Nasional Matematika berupa tes tertulis dengan bentuk soal pilihan ganda. Soal pilihan ganda dapat diskor dengan mudah, cepat, dan memiliki objektivitas yang tinggi, mengukur berbagai tingkatan kognitif, serta dapat mencakup ruang lingkup materi yang luas (Panduan Penulisan Soal Pilihan Ganda, 2007b; Lampiran Keputusan Badan Standar Nasional Pendidikan Nomor 984/BNSP/XI/2007, 2007c). Tes tertulis dengan bentuk soal pilihan ganda banyak digunakan pada ujian dengan peserta tes yang banyak.

Butir soal UN Matematika berupa tes obyektif pilihan ganda dengan empat *option* (pilihan). Model penskoran dikotomis diterapkan pada UN Matematika. Data respons butir dikotomis mempunyai dua kategori skor jawaban, yaitu jawaban benar atau tuntas (skor 1) dan jawaban salah atau tidak tuntas (skor 0) (Bond & Fox, 2007: 49; DeMars, 2010: 9). Model

penskoran dikotomus disebut model logistik dikotomus (Hambleton, Swaminathan, & Rogers, 1991: 12). Model-model logistik dikotomus dinamai sesuai dengan jumlah parameter yang dilibatkan dalam model. Model logistik satu parameter (*1-PL-model*) adalah model yang melibatkan parameter tingkat kesulitan butir. Model logistik satu parameter disebut *Rasch Model* (*RM*). Model matematika pada *1-PL-model* dan *RM* adalah *equivalent* (De Ayala, 2009: 19). Model logistik dua parameter (*2-PL-model*) adalah model yang melibatkan parameter tingkat kesulitan butir dan parameter daya beda. Model logistik tiga parameter (*3-PL-model*) adalah model yang melibatkan parameter tingkat kesulitan butir, daya beda dan parameter tebakan semu (*pseudoguessing*) (Bond & Fox, 2007: 265; Crocker & Algina, 1986: 352-354; DeMars, 2010: 9-16). Model penskoran dikotomus tersebut tidak dapat digunakan untuk membedakan kesalahan yang dilakukan oleh peserta didik, karena semua *option* yang salah diberi skor 0. Kenyataannya, kesalahan yang terjadi pada *option* yang salah dapat berasal dari penyebab kesalahan atribut yang berbeda, oleh karena itu data respons butir pada UN Matematika perlu diskor politomus.

Penyelesaian soal didasarkan pada atribut, yaitu prosedur, kompetensi, proses, atau keterampilan yang harus dimiliki peserta didik untuk menyelesaikan soal tes (Gierl, 2007a; Gierl, 2007b; Gierl, Yinggan Zheng, & Ying Cui, 2008; Robert & Gierl, 2010). Atribut dalam kategori isi (*content*) adalah materi yang digunakan di dalam kerangka tes. Atribut kategori proses (*process*) merupakan keterampilan umum. Kategori proses ini mengungkap keterampilan peserta didik yang diharapkan setelah mengikuti pelajaran dengan materi yang diberikan pada kategori isi. Atribut pada kategori keterampilan (*skill*) merupakan keterampilan proses yang spesifik.

Model penskoran politomus adalah model respons butir yang mempunyai kemungkinan jawaban lebih dari dua kategori (Wells, Hambleton, & Urip Purwono, 2008c: 1). Model penskoran meliputi *Graded Response Model* (*GRM*) dari Samejima, *Nominal Model* (*NM*) dari Bock, *Partial Credit Model* (*PCM*) dari Masters, *Generalized Partial Credit Model* (*GPCM*) dari Muraki, dan *Rating Scale Model* (*RSM*) dari Andrich (DeMars, 2010: 22; Thissen, Nelson, Rosa, et al., 2001: 143-149; Wells, Hambleton, & Urip Purwono, 2008c: 2-10; Hessen, 2009: 5-17). Model penskoran *PCM* merupakan pengembangan dari model penskoran *RM* (De Ayala, 2009: 209). Model penskoran politomus tersebut diduga dapat memberikan solusi terhadap beberapa keterbatasan model penskoran dikotomus, terkait dengan akurasi pengukuran, ketuntasan atribut yang mendasari butir soal, dan dalam penemuan informasi diagnostik yang belum didapat dari model penskoran dikotomus.

### Rumusan Masalah

Masalah dalam penelitian ini dirumuskan sebagai berikut.

1. Apa atribut-atribut yang mendasari butir-butir soal UN Matematika?
2. Bagaimana ketuntasan atribut yang mendasari butir-butir soal UN Matematika?
3. Bagaimana fungsi informasi butir dan fungsi informasi tes pada UN Matematika?
4. Bagaimana pemetaan hasil analisis menggunakan data yang diskor model *Rasch* dan model *Partial Credit*?

### Tujuan

Tujuan penelitian ini untuk menemukan:

1. atribut-atribut yang mendasari butir-butir soal UN Matematika,
2. ketuntasan atribut yang mendasari butir-butir soal UN Matematika,
3. fungsi informasi butir dan fungsi informasi tes pada UN Matematika?
4. pemetaan hasil analisis menggunakan data yang diskor model *Rasch* dan model *Partial Credit* pada UN Matematika.

### Manfaat Penelitian

Manfaat dari penelitian ini diharapkan dapat memberikan informasi tentang atribut yang mendasari butir-butir soal UN Matematika, ketuntasan atribut yang mendasari butir-butir soal UN

Matematika, fungsi informasi butir dan fungsi informasi tes pada UN Matematika, dan pemetaan hasil analisis menggunakan penskoran model *Rasch* dan penskoran model *Partial Credit*.

## **METODE PENELITIAN**

### **Jenis dan Rancangan Penelitian**

Jenis penelitian ini adalah penelitian deskriptif eksploratif terhadap data set butir soal dan respons butir peserta UN Matematika. Penelitian ini merupakan diagnosis *post-hoc* yang digambarkan sebagai pendekatan *retrofitting* (Gierl, 2007a). Pendekatan *retrofitting* dilakukan melalui analisis butir soal dan data respons butir pada UN Matematika.

### **Subjek dan Objek Penelitian**

Subjek penelitian adalah peserta UN Matematika SMP Tahun Pelajaran 2007/2008 di wilayah Kabupaten Bantul, Daerah Istimewa Yogyakarta. Objek yang dipilih adalah butir soal objektif pilihan ganda, dan respons butir peserta UN yang diskor model *Rasch* dan diskor model *Partial Credit* yang dikembangkan.

### **Teknik Pengumpulan Data**

Data berupa respons butir peserta ujian dan butir soal pada UN Matematika SMP Tahun Pelajaran 2007/2008 dikumpulkan dengan teknik dokumentasi. Data diambil dari Dinas Pendidikan dan Olahraga Daerah Istimewa Yogyakarta. Data yang dipilih butir soal UN sudah standar sehingga bias yang terjadi sudah minimal. Standarisasi butir tes UN sudah dikalibrasi sehingga dapat diperbandingkan antar seri dan antar paket dari tahun ke tahun. Butir soal UN Matematika merupakan salah satu mata pelajaran yang diujikan dalam UN, dan mata pelajaran matematika memerlukan peningkatan kualitas proses pembelajaran.

### **Teknik Analisis Data**

#### **1. Identifikasi Atribut**

Identifikasi atribut melalui pendekatan *retrofitting* yang dilakukan dengan cara mengevaluasi butir soal UN. Pendekatan *retrofitting* didasarkan pada atribut yang muncul di antara butir soal UN Matematika. Atribut yang mendasari butir soal dikategorikan menjadi atribut kategori isi (I), atribut kategori proses (P), dan atribut kategori keterampilan (K). Atribut-atribut tersebut dapat dipandang sebagai variabel kognitif. Draft hasil identifikasi atribut divalidasi oleh beberapa guru matematika SMP dan tiga pakar pendidikan matematika.

#### **2. Ketuntasan dan Ketidaktuntasan Atribut**

Untuk mendapatkan informasi tentang ketuntasan atribut dengan mengubah respons butir yang teramati menjadi peluang ketuntasan atribut.

#### **3. Model Penskoran Respons Butir**

Draft rubrik penskoran disusun melalui pendekatan *retrofitting* didasarkan pada atribut yang muncul di antara butir soal, yang dilakukan dengan cara mengevaluasi setiap *option* butir soal UN Matematika. Hasil wawancara terhadap peserta didik, dan respons peserta didik dalam menyelesaikan butir soal sebagai masukan dalam penyusunan draft model penskoran. Draft model penskoran ditelaah dan divalidasi oleh beberapa guru matematika SMP dan tiga pakar pendidikan matematika.

#### **4. Analisis Data Berdasarkan Penskoran Model *Rasch* dan Model *Partial Credit***

Untuk mendapatkan informasi tentang penguasaan atribut dengan mengubah respons butir yang teramati menjadi peluang penguasaan atau ketuntasan atribut. Estimasi kemampuan peserta dilakukan terhadap data model *Rasch* dan data model *Partial Credit*. Estimasi dilakukan menggunakan *R Programming version 2.9.0 (2009-04-17) packages irtoys, eRm (extended Rasch model)* dan *ltm (latent traits model)*, *SPSS*, dan *EXCEL*

##### **a. Analisis Data Berdasarkan Penskoran Model *Rasch***

Analisis data dikotomus difokuskan pada *I-PL model* atau *RM*. Analisis *RM* untuk menentukan: *item information fuction (IIF)*, *test information fuction (TIF)*, estimasi parameter kemampuan peserta.

b. **Analisis Data Berdasarkan Penskoran Model *Partial Credit***

Analisis data yang diskor politomus difokuskan pada model *Partial Credit* yang dikembangkan. Analisis pada model *Partial Credit* untuk menentukan: *item information fuction (IIF)*, *test information fuction (TIF)*, dan estimasi parameter kemampuan peserta didik yang ditunjukkan oleh nilai parameter *theta*.

## HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini mengambil subjek peserta Ujian Nasional Matematika SMP Tahun Pelajaran 2007/2008 di wilayah Kabupaten Bantul, Daerah Istimewa Yogyakarta. Objek yang dipilih adalah butir soal objektif pilihan ganda dan respons butir peserta UN Matematika. Data penelitian berupa butir soal dan respons butir peserta UN Matematika yang dikumpulkan dengan teknik dokumentasi. Data penelitian diambil dari Dinas Pendidikan dan Olahraga Daerah Istimewa Yogyakarta. Populasi penelitian ini adalah peserta UN Matematika meliputi 86 sekolah atau 9464 peserta di wilayah Kabupaten Bantul. Sampel diambil secara *simple area random sampling* diperoleh 12 sekolah (13,95%) atau 1016 siswa (10,74%). Berdasarkan sampel yang diperoleh untuk selanjutnya dilakukan penskoran model *Rasch* dan model *Partial Credit*.

### 1. Atribut yang Mendasari Butir Soal Ujian Nasional Matematika

Atribut yang mendasari butir soal UN dikategorikan menjadi atribut kategori isi (I), proses (P), dan keterampilan (K). Identifikasi atribut kategori isi (*content*) dibagi menjadi sub materi (a) bilangan, (b) aljabar, (c) geometri dan pengukuran, dan (d) statistika dan peluang. Sebaran hasil identifikasi atribut yang mendasari butir soal UN Matematika Tahun Pelajaran 2007/2008 disajikan pada Tabel 1.

Tabel 1  
Distribusi Atribut yang Mendasari Butir Soal Ujian Nasional  
Matematika SMP Tahun Pelajaran 2007/2008

	Bilangan	Aljabar	Geometri	Statistika	Matematika
Atribut Isi	1	2	3	3	4
Atribut Proses	6	17	20	6	36
Atr. Keterampilan	0	4	4	1	7
Jumlah	7	23	27	10	47

### 2. Ketuntasan Atribut

Untuk mendapatkan informasi tentang ketuntasan atribut dilakukan dengan mengubah respons butir yang teramati menjadi ketuntasan atribut (Gierl, 2007b). Informasi ketuntasan atribut diperoleh menurut kategori isi, proses, dan keterampilan.

a. Ketuntasan Atribut pada Model *Rasch*

1) Ketuntasan Atribut Kategori Isi pada Model *Rasch*

Ketuntasan atribut kategori isi pada butir soal matematika mempunyai rerata (0,5472). Ketuntasan atribut tertinggi dicapai pada bilangan (0,6973), diikuti aljabar (0,6040), statistika dan peluang (0,5383), dan yang terendah pada geometri dan pengukuran (0,4871). Hasil penelitian tersebut mendukung hasil penelitian yang dilakukan Tatsuoka, Corter, & Tatsuoka (2004) bahwa siswa-siswa Amerika lemah dalam hal geometri.

2) Ketuntasan Atribut Proses pada Model *Rasch*

Ketuntasan atribut kategori proses pada matematika mempunyai rerata (0,5869). Ketuntasan atribut tertinggi dicapai pada bilangan (0,7055), diikuti aljabar

(0,6), geometri dan pengukuran (0,5342), dan terendah statistika dan peluang (0,53).

3). Ketuntasan Atribut Kategori Keterampilan pada Model *Rasch*

Ketuntasan atribut kategori keterampilan mempunyai rerata (0,6548). Ketuntasan atribut tertinggi dicapai pada aljabar (0,7044), diikuti geometri dan pengukuran (0,5641), dan yang terendah pada statistika dan peluang (0,5448).

**b. Ketuntasan Atribut pada Model *Partial Credit***

1). Ketuntasan Atribut Isi pada Model *Partial Credit*

Ketuntasan atribut isi pada matematika mempunyai rerata (0,6923). Ketuntasan atribut tertinggi dicapai aljabar (0,7339), diikuti geometri dan pengukuran (0,7066), statistika dan peluang (0,7217), dan yang terendah pada butir soal bilangan (0,6985). Hasil ini mendukung penelitian yang dilakukan oleh Yi-Hsin Chen, Gorin, Thompson, et al. (2008) bahwa siswa Cina Taipei tidak pernah tuntas dalam *number sense*.

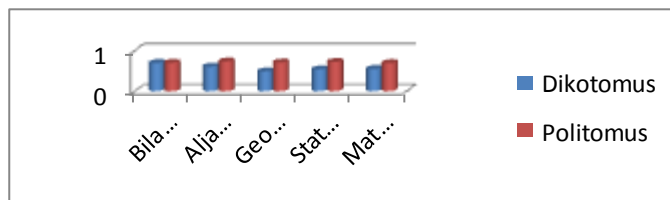
2). Ketuntasan Atribut Proses pada Model *Partial Credit*

Ketuntasan atribut isi pada matematika mempunyai rerata (0,7508). Ketuntasan atribut yang tertinggi dicapai pada aljabar (0,7723), diikuti bilangan (0,7613), geometri dan pengukuran (0,7604), dan yang terendah pada statistika dan peluang (0,7161).

3). Ketuntasan Atribut Kategori Keterampilan pada Model *Partial Credit*

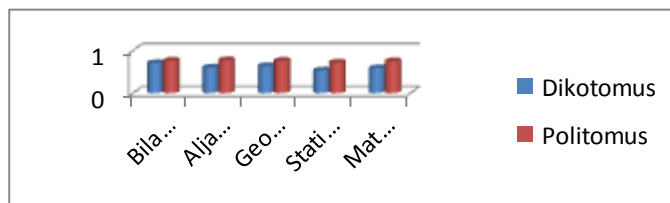
Ketuntasan atribut pada matematika mempunyai rerata (0,8104). Ketuntasan atribut yang tertinggi dicapai pada aljabar (0,8650), statistika dan peluang (0,8273), dan yang terendah pada geometri dan pengukuran (0,7102).

Untuk melihat gambaran ketuntasan atribut kategori isi secara bersama-sama disajikan pada Gambar 1, Gambar 2, dan Gambar 3.



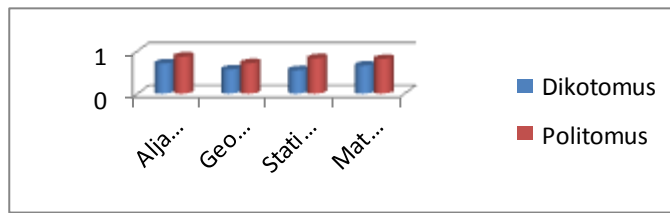
Gambar 1

*Histogram Ketuntasan Atribut Kategori Isi*



Gambar 2

*Histogram Ketuntasan Atribut Kategori Proses*



Gambar 3  
Histogram Ketuntasan Atribut Kategori Keterampilan

Jika pada UN Matematika diskor politomus, maka peluang rerata skor yang diperoleh peserta lebih tinggi dibandingkan jika diskor dikotomus. Hasil penelitian tersebut menunjukkan bahwa baik pada sub materi bilangan, aljabar, geometri dan pengukuran, statistika dan peluang, dan matematika yang diskor politomus lebih tinggi dibandingkan dengan yang diskor dikotomus.

### 3. Fungsi Informasi Butir dan Fungsi Informasi Tes

Deskripsi hasil analisis disajikan pada Tabel 2. Hasil tersebut menunjukkan bahwa data UN Matematika pada *PCM* dapat memberikan informasi tes yang lebih baik apabila dibandingkan dengan informasi tes yang dihasilkan *RM*. Demikian juga hasil analisis *PCM* dapat memberikan hasil yang lebih akurat dibandingkan *RM*. Hasil penelitian ini sejalan dengan penelitian Wasis (2009: 50) bahwa semakin tinggi kurva fungsi informasi suatu butir atau tes, maka butir atau tes tersebut semakin akurat dalam mengestimasi kemampuan peserta.

Tabel 2  
Fungsi Informasi Tes pada Model *Rasch* dan Model *Partial Credit*

Materi	<i>RM</i>		<i>PCM</i>	
	Total	<i>Std.Error</i>	Total	<i>Std.Error</i>
Bilangan	6,91	0,38042	7,80	0,35806
Aljabar	17,29	0,24049	19,55	0,22617
Geometri	18,44	0,23287	19,30	0,22763
Statistika	3,46	0,53760	3,43	0,53995
Matematika	46,10	0,14728	50,09	0,14129

### 4. Estimasi Kemampuan Siswa pada Model *Rasch* (*RM*) dan Model *Partial Credit* (*PCM*)

#### a. Estimasi Kemampuan Peserta pada Model *Rasch*

Deskripsi kemampuan peserta untuk butir soal bilangan, aljabar, geometri dan pengukuran, statistika dan peluang, dan matematika pada *RM* disajikan pada Tabel 3.

Tabel 3

Estimasi Kemampuan Peserta pada Butir Soal Bilangan, Aljabar, Geometri, Statistika, dan Matematika pada Model *Rasch*

	Rerata	Median	Simpangan Baku	<i>Std.Error</i>
Bilangan	0,40929	0,72329	1,02813	0,0420
Aljabar	0,11580	0,19543	1,06632	0,0428
Geometri	-0,42052	-0,56269	1,00634	0,0405
Statistika	-0,06200	-0,69751	0,69566	0,0324
Matematika	-0,12623	-0,21655	0,84313	0,0338

#### b. Estimasi Kemampuan Peserta pada Model *Partial Credit* (*PCM*)

Deskripsi kemampuan peserta untuk butir soal bilangan, aljabar, geometri, statistika, dan matematika pada *PCM* disajikan pada Tabel 4.

Tabel 4  
Estimasi Kemampuan Peserta pada Butir Soal Bilangan, Aljabar, Geometri, Statistika, dan  
Matematika pada Model *Partial Credit*

	Rerata	Median	Simpangan Baku	Std, Error
Bilangan	-0,06100	-0,16298	0,49308	0,0198
Aljabar	0,28985	0,27174	0,43059	0,0173
Geometri	0,23777	0,18028	0,45521	0,0183
Statistika	0,29888	0,21774	0,58764	0,0236
Matematika	0,23910	0,18807	0,34016	0,0136

## KESIMPULAN, SARAN DAN REKOMENDASI

### Kesimpulan

Berdasarkan hasil analisis data dan pembahasan diperoleh simpulan berikut.

1. Atribut yang mendasari butir soal matematika ada 47 atribut, meliputi 4 atribut isi, 36 atribut proses, dan 7 atribut keterampilan
2. Ketuntasan atribut kategori isi, proses, dan keterampilan yang mendasari butir soal pada penskoran model *Partial Credit* lebih tinggi dan lebih akurat dibandingkan dengan ketuntasan atribut pada penskoran model *Rasch*
3. Penskoran model *Partial Credit* memberikan informasi tes yang lebih baik dibandingkan dengan informasi tes yang diberikan model *Rasch*.
4. Rerata kemampuan peserta dalam matematika pada model *Partial Credit* lebih tinggi dibandingkan dengan rerata kemampuan peserta pada model *Rasch*. Informasi tes yang dihasilkan model *Partial Credit* lebih baik dibandingkan dengan informasi tes yang dihasilkan model *Rasch*.

### Saran dan Rekomendasi

1. Kepada praktisi pendidikan matematika

Informasi tentang atribut-atribut yang mendasari butir soal, ketuntasan atribut, dan pemetaan kemampuan peserta disarankan agar dapat dipertimbangkan di dalam manajemen peningkatan kualitas proses pembelajaran matematika.

2. Kepada pengambil keputusan
  - a. Penskoran model *Partial Credit* dapat menjadi solusi dari beberapa kelemahan penskoran model *Rasch*. Oleh karena itu pemangku kepentingan direkomendasikan agar penskoran pada tes objektif pilihan ganda pada matematika diarahkan pada penskoran model *Partial Credit*,
  - b. Penskoran politomus dapat menjadi solusi dari beberapa kelemahan penskoran dikotomus, oleh karena itu pada UN Matematika yang akan datang direkomendasikan untuk dikonstruksi dalam dua format, yaitu format tes objektif pilihan ganda yang disiapkan untuk diskor dikotomus, dan yang disiapkan untuk diskor politomus.
3. Kepada para peneliti

Pada penskoran *PCM* menghasilkan pengukuran yang lebih akurat dibandingkan dengan penskoran dikotomus *RM*, dan penskoran model politomus dapat memberikan solusi kelemahan penskoran model dikotomus. Para peneliti disarankan untuk melakukan penelitian yang diarahkan pada penskoran model politomus.

## DAFTAR PUSTAKA

- Bond, T. G., & Fox, C. M. (2007). *Applying the rasch model: Fundamental measurement in the human sciences*. (2<sup>nd</sup> Ed.). Mahwah: Lawrence Erlbaum Associates, Publishers.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: CBS-College Publishing.
- De Ayala, R. J.. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.
- DeMars, C. (2010). *Item Response Theory*. New York: Oxford University Press, Inc.
- Depdiknas. (2007b). *Panduan Materi SMP dan MTs Ujian Nasional Tahun Pelajaran 2007/2008, Matematika*. Pusat Penilaian Pendidikan Balitbang Depdiknas.
- \_\_\_\_\_. (2007c). *Panduan Penulisan Soal Pilihan Ganda*. Jakarta: Pusat Penilaian Pendidikan Balitbang Depdiknas.
- \_\_\_\_\_. (2007d). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20, Tahun 2007, tentang Standar Penilaian Pendidikan*.
- Gierl, M. J. (2007a). Making Diagnostic Inferences About Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method. *Journal of Educational Measurement*, 44(4), 325-340.
- \_\_\_\_\_. (6 February 2007). *Using the Attribute Hierarchy Method to Make Diagnostic Inferences about Examinees' Cognitive Skills*. Makalah disajikan dalam Association of Test Publishers Annual Meeting Palm Springs, CA.
- Gierl, M. J., Yinggan Zheng, & Ying Cui. (2008), Using the Attribute Hierarchy Method to Identify and Interpret Cognitive Skills that Produce Group Differences. *Journal of Educational Measurement*, 45 (1), 65-89.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*, Newbury Park: Sage Publications.
- Hessen, D. J. (Juni, 2009). *Applying IRT using R*. Makalah disajikan dalam Pelatihan Aplikasi Program R dalam Bidang Psikometri dan Penilaian Pendidikan, di PPs Universitas Negeri Yogyakarta.
- Roberts, M. R., & Gierl, M. J. (2010). Developing Score Reports for Cognitive Diagnostic Assessment. *Educational Measurement: Issues and Practice*. 29 (3), 25-38.
- Tatsuoka, K., Corter, J.E., & Tatsuoka, C. (2004). Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R A cross a Sample of 20 Countries. *American Educational Research Journal*, 41(4), 901-926.
- Thissen, D., Nelson, L., Rosa, K., et al. (2001). Item Response Theory for Items Scored in More Than Two Categories. Dalam D. Thissen & H. Wainer. *Test Scoring* (pp. 141-184). New Jersey: Lawrence Erlbaum Associates Publishers.



- Wasis. (2009). *Penskoran Model Partial Credit Pada Item Multiple True-False Bidang Fisika*. Disertasi doktor, tidak diterbitkan. Yogyakarta: Universitas Negeri Yogyakarta.
- Wells, C. S., Hambleton, R.K., & Urip Purwono. (18-24 Juni 2008a). *Assessing the Fix of IRT Models to Item Response Data*. Makalah disajikan dalam Pelatihan Asesmen Pendidikan dan Psikologi (Psikometri), di PPs Universitas Negeri Yogyakarta.
- Yi-Hsin Chen, Gorin, J.D., Thompson, M.S., et al. (2008). An alternative examination of Chinese Taipei mathematics achievement: Application of the rule-space method TIMSS 1999 data. *IERI Monograph Series, Issues and Methodologies in Large-Scale Assessments*. 1, 23-49.

