

PERBANDINGAN PENYEKORAN MODEL *RASCH* DAN MODEL *PARTIAL CREDIT* PADA MATEMATIKA

Awal Isgiyanto

FKIP Universitas Bengkulu
email: awalunib@yahoo.co.id

Abstrak

Penelitian ini bertujuan untuk menentukan atribut butir soal; kemampuan peserta menjawab butir soal; penyekoran model *Rasch* dan *Partial Credit*; serta nilai fungsi informasi model *Rasch* dan *Partial Credit* tes matematika. Metode penelitiannya adalah *retrofitting* dengan subjek peserta ujian nasional matematika SMP di Kabupaten Bantul. Objeknya berupa butir dan respons butir peserta. Hasilnya adalah *pertama*, atribut yang mendasari butir soal ada 47 yang terdiri atas empat atribut isi, 36 atribut proses, dan tujuh atribut keterampilan; *kedua*, kemampuan tertinggi peserta ada pada butir soal statistika dan peluang, diikuti aljabar dan geometri, sedangkan kemampuan terendah pada butir soal bilangan; *ketiga*, tingkat kesulitan tertinggi model *Rasch* terletak pada bilangan, aljabar, serta statistika dan peluang dan terendah pada geometri dan pengukuran, sedangkan tertinggi pada model *Partial Credit* terletak pada *threshold 2*, diikuti *threshold 1*, dan terendah pada *threshold 3*; serta *keempat*, nilai fungsi informasi model *Partial Credit* lebih baik dan akurat daripada model *Rasch*.

Kata kunci: penyekoran model *rasch*, model *partial credit*

COMPARISON OF SCORING OF THE *RASCH* MODEL AND THE *PARTIAL CREDIT* MODEL IN MATHEMATICS

Abstract

The study is aimed at determining the test item attributes, testees' ability in answering the test items, model scoring, and the degree of the information function of the *Rasch* model and the *Partial Credit* model of mathematics tests. The research method is that of *retrofitting* involving as subjects participants of the junior-high-school mathematics national examination of Bantul Regency. The research object consists of the participants' items and item responses. Findings show the following. First, there are 47 attributes underlying the test items consisting of four attributions, 36 process attributions, and seven skill attributions. Second, the highest participants' ability is found on the statistical and chance test items, followed by algebra and geometry, while the lowest ability is on math test items. Third, the highest difficulty level of the *Rasch* model is found on math, algebra, and statistical and chance test items, while the lowest difficulty level is on geometry and measurement. Meanwhile, the highest level on the *Partial Credit* model is on *threshold 2*, followed by *threshold 1*, and the lowest is on *threshold 3*. Fourth, the value of the information function of the *Partial Credit* model is better and more accurate than that of the *Rasch* model.

Keywords: scoring *rasch* model, *partial credit* model

PENDAHULUAN

Kalibrasi dalam *Item Response Theory* (*IRT*) merupakan proses penentuan parameter

respons suatu butir, dan fungsi responsnya memuat parameter butir dan parameter person. Tujuan kalibrasi adalah untuk mendapatkan

nilai estimasi parameter butir dan parameter person. Parameter butir dapat berupa tingkat kesulitan butir dan parameter person dapat berupa kemampuan peserta.

Nilai fungsi informasi tes (*test information function-TIF*) merupakan penjumlahan seluruh nilai fungsi informasi butir. Fungsi informasi butir (*item information function-IIF*) merupakan suatu fungsi untuk menjelaskan kekuatan suatu butir, pemilihan butir, dan perbandingan beberapa perangkat tes. Melalui *IIF* dapat diketahui butir tes yang cocok dengan model, sehingga dapat digunakan untuk seleksi butir tes.

Untuk menyelesaikan butir soal matematika diperlukan sejumlah tahapan penyelesaian. Pada setiap tahapan diperlukan penguasaan atribut yang mendasari butir soal tersebut. Atribut didefinisikan sebagai prosedur, proses, keterampilan, atau kompetensi yang harus dimiliki peserta untuk menyelesaikan butir soal (Gierl, 2007a; Gierl, Yinggan Zheng, & Ying Cui, 2008; Robert & Gierl, 2010). Atribut diperlukan untuk menyelesaikan butir soal belum tentu dikuasai dan diterapkan oleh peserta dengan tepat. Pada Penyekoran dikotomus, hanya peserta yang sudah menguasai dan menerapkan dengan tepat semua atribut matematika yang diperlukan untuk menjawab dengan betul saja yang dihargai. Peserta yang belum menguasai dan menerapkan dengan tepat seluruh atribut tidak diberi *credit*. Oleh karena itu perlu inovasi model Penyekoran dengan mengembangkan model Penyekoran politomus pada butir soal objektif pilihan ganda pada Matematika.

Respons butir dikotomus mempunyai dua kategori skor jawaban, yaitu jawaban betul (skor 1) dan jawaban salah (skor 0) (Bond & Fox, 2007: 49; DeMars, 2010: 9). Model Penyekoran dikotomus disebut model logistik dikotomus. Model logistik dikotomus dinamai sesuai dengan banyaknya parameter yang dilibatkan dalam model (Hambleton, Swaminathan, & Rogers, 1991: 12). Model

logistik satu parameter (*1-PL-model*) atau *Rasch Model (RM)* adalah model yang melibatkan parameter tingkat kesulitan butir.

Penyekoran politomus adalah model respons butir yang mempunyai kemungkinan jawaban lebih dari dua kategori (Wells, Hambleton, & Urip Purwono, 2008: 1). Model Penyekoran politomus 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, 2008: 2-10; Hessen, 2009: 5-17). *Partial Credit Model (PCM)* dikembangkan dari *Rasch Model (RM)* oleh Masters. *RM* untuk data skor dikotomus dikembangkan menjadi skor lebih dari dua kategori (politomus). *PCM* tidak mensyaratkan langkah penyelesaian butir tes harus berurutan dan tidak harus mempunyai kesulitan yang sama (De Ayala, 1993). *Threshold* pada Penyekoran *PCM* dari suatu kategori ke kategori berikutnya tidak selalu lebih besar. *PCM* juga cocok untuk butir yang diskor dalam kategori berjenjang, akan tetapi indeks kesukaran dalam setiap langkah tidak perlu terurut. Skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih tinggi dari pada skor kategori yang lebih rendah. Respons butir pada penelitian ini dapat betul pada langkah tertentu, tetapi dapat salah pada langkah yang lain. Pada kasus respons seperti itu, disarankan untuk menggunakan Penyekoran *PCM*. Pada *RM* atau *PCM* diasumsikan bahwa parameter tingkat kesulitan butir merupakan satu-satunya karakteristik butir yang mempengaruhi kinerja peserta.

Pada Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 22 Tahun 2006 Tentang Standar Isi untuk Satuan Pendidikan Dasar dan Menengah

disebutkan bahwa ruang lingkup mata pelajaran Matematika pada satuan pendidikan SMP/MTs meliputi aspek-aspek (1) bilangan, (2) aljabar, (3) geometri dan pengukuran, dan (4) statistika dan peluang. Pada tingkat SMP, matematika berfungsi untuk mengembangkan kemampuan menghitung, mengukur, merumuskan dan menggunakan rumus matematika yang diperlukan dalam kehidupan sehari-hari melalui materi bilangan, aljabar, geometri dan pengukuran, dan statistika dan peluang. Demikian pentingnya rekomendasi kepada pemangku kepentingan, maka perlu penelitian tentang fungsi informasi tes model *Rasch* dan model *Partial Credit* pada Matematika.

Berdasarkan uraian pada latar belakang, rumusan masalah penelitian adalah “bagaimanakah nilai fungsi informasi tes model *Rasch* dan model *Partial Credit* pada tes objektif pilihan ganda pada matematika?”. Tujuan penelitian ini untuk menemukan: (1) atribut yang mendasari butir soal, (2) model Penyekoran dikotomis model *Rasch* dan politomis model *Partial Credit*, dan (3) nilai fungsi informasi tes model *Rasch* dan model *Partial Credit* pada tes objektif pilihan ganda pada matematika.

METODE

Penelitian ini menggunakan pendekatan *retrofitting* (Gierl, 2007a). Pendekatan *retrofitting* dilakukan melalui analisis butir soal dan data respons butir pada UN Matematika. Subjek penelitian adalah peserta UN Matematika SMP Tahun Pelajaran 2007/2008 di wilayah Kabupaten Bantul, Daerah Istimewa Yogyakarta. Objek yang dipilih adalah 40 butir soal dan 1016 respons butir peserta UN Matematika. Sampel diambil secara *simple random sampling* diperoleh 12 sekolah atau 1016 peserta. Data *ex post facto* 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.

Identifikasi atribut dilakukan melalui pendekatan *retrofitting* dengan cara menganalisis data berupa butir soal UN Matematika. Pendekatan *retrofitting* didasarkan pada atribut yang muncul di antara butir soal. Konstruksi suatu model kognitif diidentifikasi untuk menemukan atribut yang mendasari butir soal melalui analisis tugas pada butir soal yang mewakili suatu *domain* tertentu. Atribut yang mendasari butir soal sejalan dengan kompetensi dasar (KD). Atribut yang mendasari butir soal dikategorikan menjadi atribut kategori isi (I), atribut kategori proses (P), dan atribut kategori keterampilan (K). Hasil identifikasi atribut divalidasi oleh siswa, guru matematika SMP, dan pakar pendidikan matematika.

Data berupa butir soal UN Matematika dianalisis untuk mendapatkan Rubrik Penyekoran. Rubrik Penyekoran disusun melalui pendekatan *retrofitting* didasarkan pada atribut yang muncul di antara butir soal. Untuk memperoleh *option* yang betul dapat dilakukan melalui prosedur yang bermacam-macam. Peserta memilih *option* yang betul diberi skor maksimum (skor 3). Pada tiga *option* yang salah (*distraktor*) dikategorikan menurut bobot kesalahannya menjadi kategori ringan (skor 2), kategori sedang (skor 1), dan kategori berat (skor 0). Rubrik Penyekoran divalidasi oleh siswa, guru matematika SMP, dan pakar pendidikan matematika. Validitas Rubrik Penyekoran didasarkan pada hasil validasi. Kesahihan skor yang diperoleh juga didasarkan pada validitas Rubrik Penyekoran tersebut.

Analisis menggunakan *R Programming version 2.9.0 (2009-04-17) packages irtoys*, dan *eRm (extended Rasch model)*. Analisis data juga menggunakan bantuan *SPSS* dan *EXCEL*. Analisis data dikotomis difokuskan pada model *Rasch*. Analisis terhadap data model *Rasch* untuk menemukan harga parameter tingkat kesulitan butir (*difficulty*) dan *test*

information function (TIF). Analisis terhadap data yang diskor politomus difokuskan pada model *Partial Credit*. Analisis data pada Penyekoran model *Partial Credit* untuk menemukan parameter tingkat kesulitan butir (*thresholds*), *test information function (TIF)*, dan kemampuan peserta.

HASIL PENELITIAN DAN PEMBAHASAN Atribut yang Mendasari Butir Soal

Atribut yang menyusun konstruk suatu model kognitif diidentifikasi melalui analisis tugas pada butir soal yang mewakili suatu *domain* tertentu. Untuk mengidentifikasi atribut yang mendasari butir soal UN Matematika digunakan pendekatan diagnosis *post-hoc*, yang digambarkan sebagai pendekatan *retrofitting* (Gierl, 2007b). Pendekatan ini dilakukan dengan cara menganalisis butir soal UN Matematika. Pendekatan *retrofitting* tersebut didasarkan pada atribut yang muncul di antara butir soal UN Matematika.

Sebaran hasil identifikasi atribut yang mendasari butir soal disajikan pada Tabel 1. Pada Tabel 1 tampak bahwa ada 47 atribut yang mendasari butir-butir soal UN Matematika, meliputi 4 atribut kategori isi, 36 atribut kategori proses, dan 7 atribut kategori keterampilan.

Tiga puluh enam atribut kategori proses tersebut ada tiga atribut yang tidak termasuk KD. Atribut yang dimaksud adalah atribut (At.P.1) kompetensi melakukan proses pemikiran logis, atribut (At.P.2) kompetensi menentukan nilai bentuk aljabar, dan atribut (At.P.3) kompetensi melakukan konversi satuan panjang. Atribut (At.P.1) tersebut

teridentifikasi pada pemecahan masalah butir soal 2, 3, 5, 6, 7, 8, 14, 15, 32, 33, 35, dan 39. Atribut (At.P.2) teridentifikasi pada butir soal 19. Atribut (At.P.3) teridentifikasi pada butir soal 32 dan 33. Ketiga atribut tersebut tidak termasuk KD, akan tetapi merupakan kompetensi sangat penting dalam penyelesaian butir soal UN Matematika Tahun Pelajaran 2007/2008. Untuk selanjutnya, ketiga atribut tersebut dapat dipertimbangkan dalam revisi KD mata pelajaran matematika SMP.

Rubrik dan Model Penyekoran

Penyusunan Rubrik Penyekoran dilakukan melalui pendekatan *retrofitting*. Pendekatan *retrofitting* didasarkan pada atribut yang muncul di antara butir soal. Konstruk suatu model kognitif setiap distraktor diidentifikasi untuk menemukan atribut yang mendasarinya. Identifikasi dilakukan melalui analisis tugas pada butir soal yang mewakili suatu *domain* tertentu.

Untuk menjawab butir soal matematika diperlukan sejumlah tahapan penyelesaian. Pada setiap tahapan diperlukan penguasaan atribut yang mendasari butir soal yang bersangkutan. Oleh karena itu perlu inovasi model Penyekoran dengan mengembangkan model Penyekoran politomus pada tes objektif pilihan ganda pada Matematika.

Rubrik Penyekoran data UN Matematika pada penelitian ini difokuskan pada paradigma pengujian unidimensi. Hasil penelitian Walker & Beretvas (2003) menyatakan bahwa menurut model unidimensi, dimensi tunggal dapat mewakili

Tabel 1. Distribusi Atribut yang Mendasari Butir Soal

	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

beberapa kombinasi kemampuan matematika umum dan komunikasi matematika. UN merupakan tes pendidikan berskala besar yang berfungsi untuk mengukur dan menyusun peringkat peserta berdasarkan pada sifat laten unidimensional.

Penyekoran Rasch Model (RM)

Penyekoran *RM* mempunyai dua kategori. Pada Rubrik Penyekoran dikotomus *RM*, *option* yang betul diberi skor "1", dan tiga *option* yang salah diberi skor "0". Peserta yang diberi skor "0", sesungguhnya peserta yang bersangkutan melakukan kesalahan yang berbeda. Jika dalam UN Matematika masih menggunakan tes objektif pilihan ganda, maka perlu alternatif Penyekoran politomus agar dapat digunakan untuk mengontrol distraktor dan untuk mengungkap informasi diagnostik yang belum didapat dari model Penyekoran dikotomus.

Ketuntasan atribut setiap *option* pada kasus butir soal 32 tampak bahwa peserta yang memilih *option* A, C, dan D diberi skor "0". Padahal peserta yang diberi skor "0" tersebut sudah menguasai atribut yang mendasari butir soal yang dikembangkan dari kompetensi dasar (KD), yaitu (At.I.3) konsep dasar geometri, (At.P.27) kompetensi mengidentifikasi sifat-sifat kubus, balok, prisma dan limas serta bagian-bagiannya, (At.I.1) konsep dasar bilangan, dan atribut (At.P.4) kompetensi melakukan operasi hitung bilangan bulat dan pecahan. Penyekoran model dikotomus pada kasus butir soal 32 tersebut kurang tepat dan kurang adil, karena peserta sudah menerapkan dengan benar semua KD pada Matematika SMP, dan peserta hanya melakukan kesalahan dalam mengkonversikan satuan panjang, yang sesungguhnya bukan KD yang harus dikuasai oleh siswa SMP.

Pada kasus butir soal 33 tampak bahwa peserta yang memilih *option* B diberi skor "1", dan peserta yang memilih *option* A, C dan D diberi skor "0". Pada Rubrik

Penyekoran tampak bahwa peserta yang memilih *option* A melakukan kesalahan dalam menentukan rumus luas permukaan kotak kayu $L=pl+lt+pt$, yang berarti peserta tidak tuntas atribut (A1).P.1. Kompetensi melakukan proses pemikiran logis, (A2).I.1. Konsep dasar geometri dan pengukuran: balok, dan (A3).P.2. Kompetensi menghitung luas permukaan dan volume kubus, balok, prisma dan limas. Peserta yang memilih *option* C melakukan kesalahan dalam menentukan rumus luas permukaan kotak kayu $L= pl+lt+pt$, dan dalam melakukan konversi satuan panjang $14.000 \text{ cm}^2 = 14 \text{ m}^2$, yang berarti peserta tidak tuntas atribut (A1).P.1. Kompetensi melakukan proses pemikiran logis, (A2).I.1. Konsep dasar geometri dan pengukuran: balok, (A3).P.2. Kompetensi menghitung luas permukaan dan volume kubus, balok, prisma dan limas, dan (A6).P.4. Kompetensi melakukan konversi satuan panjang. Peserta yang memilih *option* D hanya melakukan kesalahan melakukan konversi satuan panjang $28.000 \text{ cm}^2 = 28 \text{ m}^2$, yang berarti peserta tidak tuntas pada atribut (A6).P.4. Kompetensi melakukan konversi satuan panjang saja.

Penyekoran Model Politomus

Butir soal UN Matematika tidak dikonstruksi untuk dilakukan Penyekoran politomus. Penyusunan rubrik Penyekoran politomus dilakukan melalui pendekatan *retrofitting*. Respons butir pada Penyekoran politomus mempunyai kemungkinan jawaban lebih dari dua kategori. Untuk memperoleh jawaban yang betul, peserta dapat menyelesaikan butir soal melalui tahapan yang berbeda-beda. Oleh karena itu peserta yang memilih *option* yang betul diberi skor maksimum (skor 3). Proses Penyekoran pada *option* yang salah dilakukan secara *partial*, untuk selanjutnya dikategorikan menurut bobot kesalahannya menjadi kategori ringan (skor 2), kategori sedang (skor 1), dan kategori berat (skor 0).

Penyekoran politomus *PCM* yang dikembangkan bertujuan untuk mengatasi kelemahan model Penyekoran dikotomus, yaitu tidak dapat digunakan untuk mengontrol distraktor dan untuk mengetahui kesalahan yang dilakukan oleh peserta, karena semua *option* yang salah diberi skor "0". Kenyataannya, kesalahan pada distraktor dapat berasal dari sumber kesalahan yang berbeda. Pengembangan model Penyekoran politomus ini dengan menyesuaikan *PCM* dari Masters dengan pertimbangan (1) model Penyekoran politomus dapat memberikan *credit* pada penerapan atribut dengan benar, (2) model Penyekoran politomus akan memberikan peluang meningkatkan akurasi pengukuran dibandingkan dengan model dikotomus, (3) antar tahapan dalam menyelesaikan butir soal matematika tidak selalu mempunyai tingkat kesulitan yang sama, dan (4) kategori yang lebih tinggi tidak selalu mempunyai *threshold* yang lebih tinggi, demikian juga kategori yang lebih rendah tidak selalu mempunyai *threshold* yang lebih rendah.

Tingkat Kesulitan Butir

Pada penelitian ini difokuskan pada Penyekoran dikotomus model *Rasch (RM)* dan Penyekoran politomus model *Partial Credit (PCM)*. *PCM* merupakan perluasan dari *RM*, sehingga diasumsikan bahwa semua butir soal mempunyai daya beda yang sama. Pada *RM* atau *PCM* diasumsikan bahwa parameter tingkat kesulitan butir merupakan satu-satunya karakteristik butir yang mempengaruhi kinerja peserta didik, yang berarti parameter tebakan semu © diasumsikan tidak berpengaruh terhadap peluang kemampuan peserta untuk menjawab butir soal dengan betul. Peluang peserta didik berhasil mengerjakan butir soal tergantung pada kemampuan (*ability*) dan tingkat kesulitan (*difficulty*) butir soal yang dikerjakannya.

Kemampuan peserta pada matematika dinyatakan oleh parameter *theta*. Estimasi kemampuan peserta pada butir soal matematika

merupakan gabungan dari butir soal bilangan, aljabar, geometri dan pengukuran, dan butir soal statistika dan peluang. Kemampuan peserta pada butir soal statistika dan peluang 0,29888, kemampuan peserta pada butir soal aljabar 0,28985, kemampuan peserta pada butir soal geometri 0,23777, dan kemampuan peserta pada butir soal bilangan -0,06100. Deskripsi kemampuan peserta pada UN mata pelajaran matematika diperoleh rerata 0,23910, median 0,18807, simpangan baku 0,34016, dengan *standard error of estimation* 0,0136.

Tingkat Kesulitan Butir pada Model *Rasch (RM)*

Deskripsi tingkat kesulitan butir *RM* pada butir soal UN Matematika mempunyai rerata -0,085, median -0,1576, *range* 2,77, minimum -1,51, maksimum 1,26, dan *standard error of estimation* 0,1222. Rerata tingkat kesulitan tertinggi pada bilangan 0,540719, kemudian pada aljabar 0,054208, pada ststistika dan peluang -0,33709, dan terendah pada geometrid dan pengukuran -0,33709. Tingkat kesulitan sebesar -0,085 mempunyai makna bahwa peserta diharapkan dapat menyelesaikan butir soal dengan benar jika mempunyai kemampuan minimal -0,085. Parameter tingkat kesulitan butir merupakan parameter lokasi yang menunjukkan posisi kurva karakteristik butir dalam hubungannya dengan skala kemampuan. Parameter tingkat kesulitan butir digambarkan oleh suatu titik pada skala kemampuan dimana peluang menjawab betul sebesar 0,5. Semakin besar nilai parameter beta, maka semakin besar kemampuan yang diperlukan peserta untuk mendapatkan peluang 50% menjawab butir soal dengan betul.

Tingkat Kesulitan Butir pada Model *Partial Credit (PCM)*

PCM tidak mensyaratkan langkah penyelesaian butir tes harus berurutan dan tidak harus mempunyai kesulitan yang sama

(De Ayala, 1993). *PCM* yang dikembangkan ini mempunyai empat kategori, sehingga analisis *PCM* menghasilkan tiga *thresholds* (tingkat kesulitan) untuk setiap butir, yakni *Thresholds* Butir Soal Bilangan, *Thresholds* Butir Soal Aljabar, *Thresholds* Butir Soal Geometri dan Pengukuran, dan *Thresholds* Butir Soal Statistika dan Peluang.

Berdasarkan *Thresholds* Butir Soal Bilangan diperoleh hasil dari *threshold 1* ke *threshold 2* yang mengalami kenaikan pada butir soal 1 dan 3 serta yang mengalami penurunan pada butir soal 2, 4, 9, dan 10. *Threshold 2* ke *threshold 3* yang mengalami kenaikan pada butir soal 2 dan 10, dan yang mengalami penurunan pada butir soal 1, 3, 4, dan 9.

Hasil analisis *Thresholds* Butir Soal Aljabar dari *threshold 1* ke *threshold 2* yang mengalami kenaikan pada butir soal 6, 8, 11, 12, 13, dan 15, dan yang mengalami penurunan pada butir soal 5, 7, 14, dan 16. *Threshold 2* ke *threshold 3* yang mengalami kenaikan hanya pada butir soal 14, dan pada butir soal aljabar yang lain mengalami penurunan.

Hasil analisis *Thresholds* Butir Soal Geometri dan Pengukuran Dari *threshold 1* ke *threshold 2* yang mengalami kenaikan pada 11 butir soal, yaitu butir soal 22, 23, 24, 27, 28, 30, 31, 33, 34, dan 36, dan yang mengalami penurunan pada 5 butir soal, yaitu butir 25, 9, 32, 35, dan 37. *Threshold 2* ke *threshold 3* yang mengalami kenaikan pada 3 butir soal, yaitu butir 25, 31, dan 37, dan yang mengalami penurunan pada 13 butir soal yang lain.

Hasil analisis *Thresholds* Butir Soal Statistika dan Peluang Dari *threshold 1* ke *threshold 2* yang mengalami kenaikan butir soal 38 dan 39, yang mengalami penurunan pada butir soal 40 terjadi pada *threshold 2* ke *threshold 3*.

Penyekoran politomus memiliki tingkat kesulitan (*threshold*) lebih dari 1. *PCM* dengan 4 kategori Penyekoran 0, 1, 2, dan

3, maka butir soal tersebut mempunyai 3 *threshold*. *Threshold 1* dari kategori 0 ke kategori 1, *Threshold 2* dari kategori 1 ke kategori 2, dan *Threshold 3* dari kategori 2 ke kategori 3. Perpotongan grafik antara kategori 0 dan kategori 1 menghasilkan suatu titik. Jika dari titik tersebut ditarik garis lurus memotong sumbu *ability*, maka titik potong dengan sumbu *ability* disebut *threshold 1*. Perpotongan grafik antara kategori 1 dan kategori 2 menghasilkan suatu titik. Jika dari titik tersebut ditarik garis lurus memotong sumbu *ability*, maka titik potong dengan sumbu *ability* disebut *threshold 2*. Perpotongan grafik antara kategori 2 dan kategori 3 menghasilkan suatu titik. Jika dari titik tersebut ditarik garis lurus memotong sumbu *ability*, maka titik potong dengan sumbu *ability* disebut *threshold 3*.

Butir soal 10 mempunyai *threshold 1* sebesar 3,61417, mempunyai makna bahwa peserta dapat menyelesaikan dengan betul dari kategori 0 ke kategori 1 apabila peserta mempunyai kemampuan minimal 3,61417. Nilai *threshold 2* sebesar -2,77800, mempunyai makna bahwa para peserta harus memiliki kemampuan minimal sebesar -2,77800 untuk dapat menyelesaikan dengan betul dari kategori 1 ke kategori 2. Nilai *threshold 3* sebesar -1,42640, mempunyai makna bahwa peserta harus memiliki kemampuan minimal sebesar -1,42640, untuk dapat menyelesaikan kategori 2 dan kategori 3 dengan betul.

Untuk melihat kecenderungan *Threshold 1*, *Threshold 2*, dan *Threshold 3* pada bilangan, aljabar, geometri, statistika, dan matematika disajikan pada Gambar 1. Pada Gambar 1 tampak bahwa *threshold* pada submateri bilangan semakin tinggi posisi *threshold*, maka nilai *threshold* semakin rendah, yang berarti semakin mudah. Pada submateri aljabar, geometri, dan statistika tampak bahwa dari *threshold 1* ke *threshold 2* semakin sulit, dan dari *threshold 2* ke *threshold 3* semakin mudah.

Butir soal UN Matematika secara keseluruhan mempunyai rerata *threshold 1* sebesar 0,07335, rerata *threshold 2* sebesar 0,37047, dan rerata *threshold 3* sebesar -0,99802. Berdasarkan hasil analisis tersebut dapat dikatakan bahwa tingkat kesulitan pada butir soal UN Matematika yang paling tinggi pada umumnya adalah *threshold 2*, diikuti *threshold 1*, dan yang paling mudah pada umumnya terjadi pada *threshold 3*.

Fungsi Informasi Tes

Nilai fungsi informasi tes yang diberikan pada *RM* dan *PCM* pada bilangan, aljabar, geometri dan pengukuran, statistika dan peluang, dan matematika disajikan pada Tabel 2.

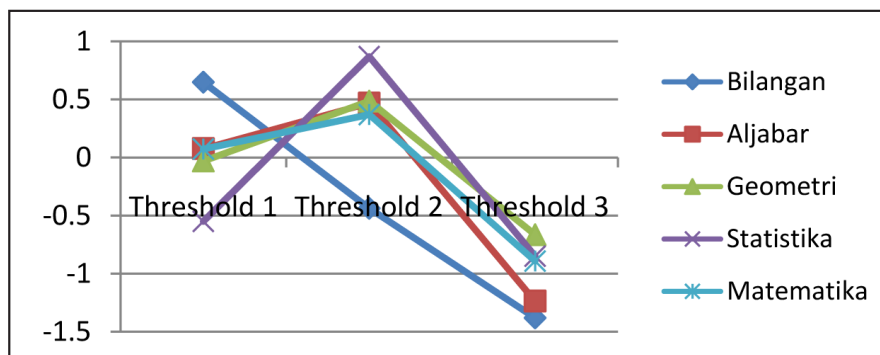
Pada Tabel 2 tampak bahwa nilai fungsi informasi tes pada *PCM* sebesar 50,09 lebih besar dari pada estimasi skor fungsi informasi tes pada *RM* sebesar 46,1. Hasil estimasi menunjukkan bahwa data UN Matematika pada *PCM* dapat memberikan nilai fungsi informasi tes yang lebih baik dan lebih akurat dibandingkan dengan nilai fungsi informasi tes yang dihasilkan *RM*.

Hasil *TIF* yang diperoleh pada *RM* dan *PCM* submateri bilangan, aljabar, geometri dan pengukuran, statistika dan peluang, dan matematika disajikan pada Tabel 2. Pada Tabel 2 tampak bahwa nilai fungsi informasi tes *RM* pada matematika lebih

kecil dibandingkan nilai fungsi informasi tes pada *PCM*. Hal ini bisa terjadi karena model Penyekoran politomus mempunyai kategori lebih banyak dibandingkan dengan kategori pada model dikotomus, sehingga varians pada model politomus lebih kecil dibandingkan dengan varians pada model dikotomus. Varians pada model politomus lebih kecil dibandingkan varians pada model dikotomus, sehingga nilai fungsi informasinya model politomus lebih besar dibandingkan dengan nilai fungsi informasi pada dikotomus. Kurva fungsi informasi model politomus lebih tinggi dan lebih sempit, sehingga hasil estimasinya lebih akurat. Hasil tersebut menunjukkan bahwa data UN Matematika pada *PCM* dapat memberikan nilai fungsi informasi tes yang lebih baik dibandingkan dengan nilai fungsi informasi tes yang dihasilkan *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 baik dalam mengestimasi kemampuan peserta.

SIMPULAN

Berdasarkan hasil dan pembahasan, dapat ditarik beberapa simpulan. *Pertama*, atribut yang mendasari butir soal pada matematika ada 47 atribut yang terdiri atas empat atribut isi, 36 atribut proses, dan tujuh atribut keterampilan. *Kedua*, kemampuan



Gambar 1. Diagram Garis *Threshold 1*, *Threshold 2*, dan *Threshold 3* pada Submateri Bilangan, Aljabar, Geometri, Statistika, dan Matematika

Tabel 2. Fungsi Informasi Tes untuk Model *Rasch* dan Model *Partial Credit*

Submateri	RM		PCM	
	Total	Std.Error	Total	Std.Error
Bilangan	6,91	0,38042	7,8	0,35806
Aljabar	17,29	0,24049	19,55	0,22617
Geometri	18,44	0,23287	19,3	0,22763
Statistika	3,43	0,53995	3,46	0,53760
Matematika	46,1	0,14728	50,09	0,14129

peserta yang tertinggi pada butir soal statistika dan peluang, diikuti kemampuan peserta pada butir soal aljabar, kemampuan peserta pada butir soal geometri, dan terendah dan kemampuan peserta pada butir soal bilangan. *Ketiga*, tingkat kesulitan tertinggi pada model *Rasch* dicapai pada bilangan, kemudian aljabar, ststistika dan peluang, dan terendah pada geometri dan pengukuran. *Threshold* model *Partial Credit* pada butir soal matematika yang tertinggi adalah *threshold 2*, diikuti *threshold 1*, dan yang terendah terjadi pada *threshold 3*. *Keempat*, hasil fungsi informasi tes model *Partial Credit* dapat memberikan nilai fungsi informasi tes yang lebih baik dan lebih akurat dibandingkan dengan nilai fungsi informasi tes yang dihasilkan model *Rasch*.

DAFTAR PUSTAKA

- Bond, T. G., & Fox, C. M. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. (2nd Ed.). Mahwah: Lawrence Erlbaum Associates, Publishers.
- De Ayala, R. J. 1993. "Methods, Plainly Speaking: An Introduction to Polytomous Item Response Theory Models". *Measurement and Evaluation in Counseling and Development*, 25, 172-189
- DeMars, C. 2010. *Item Response Theory*. New York: Oxford University Press, Inc.
- Depdiknas. 2006. *Peraturan Menteri Pendidikan Nasional Nomor 22, Tahun 2006, tentang Standar Isi untuk Satuan Pendidikan Dasar dan Menengah*.
- 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.
- Gierl, M. J. 2007b. "Using the Attribute Hierarchy method to make diagnostic inferences about examinees' cognitive skills". *Makalah* disajikan dalam Association of Test Publishers Annual Meeting Palm Springs, 6 February 2007.
- 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. 2009. "Applying IRT using R". *Makalah* disajikan dalam Pelatihan Aplikasi Program R dalam Bidang Psikometri dan Penilaian Pendidikan, di PPs Universitas Negeri Yogyakarta, Juni, 2009.

- Roberts, M. R., & Gierl, M. J. 2010. "Developing Score Reports for Cognitive Diagnostic Assessment". *Educational Measurement: Issues and Practice*. 29 (3), 25-38.
- 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.
- Walker, C. M., & Beretuas, S. N. 2003. Comparing Multidimensional and Undimensional Proficiency Classifications: Multidimensional IRT as a Diagnostic Aid. *Journal of Educational Measurement*, 40, 255-275.
- Wasis. 2009. "Penyekoran 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. "Polytomous Response IRT Models and Applications". *Makalah* Disajikan dalam Pelatihan Asesmen Pendidikan dan Psikologi (Psikometri), 18-24 Juni 2008 di PPs Universitas Negeri Yogyakarta.