

## **PENDETEKSIAN KEBERFUNGSIAN BUTIR PEMBEDA DENGAN INDEKS VOLUME SEDERHANA BERDASARKAN TEORI RESPONS BUTIR MULTIDIMENSI**

*Heri Retnawati*

Fakultas Matematika dan Ilmu Pengetahuan Alam UNY  
retnawati\_heriuny@yahoo.co.id

### **Abstrak**

Studi ini bertujuan untuk mengidentifikasi keberfungsian butir pembeda (DIF) berdasarkan teori respons butir dengan model dua dimensi. Penelitian ini merupakan penelitian eksploratif dengan pendekatan kuantitatif. Data yang digunakan pada penelitian ini adalah data ujian nasional mata pelajaran matematika 2003 yang sudah terbukti memuat 2 dimensi (Badrun Kartowagiran & Heri Retnawati, 2007). Untuk mendeteksi DIF, digunakan indeks volume sederhana yang merupakan pengembangan dari indeks daerah bertanda. Pada penelitian ini peserta tes dikelompokkan berdasarkan jenis kelaminnya, kelompok siswa laki-laki dan siswa perempuan. Estimasi parameter butir dilakukan dengan bantuan BILOG dan TESTFACT, kemudian estimasi volume bertanda dan menggambar permukaan karakteristik butir dengan program MAPLE. Hasil analisis menunjukkan bahwa dari 28 butir yang dianalisis, sebanyak 2 butir berpotensi memuat DIF seragam dan 26 butir berpotensi memuat DIF yang tidak seragam.

**Kata kunci:** *teori respons butir unidimensi, teori respons butir multidimensi DIF, UPD, SPD*

## **INDENTIFYNG DIFFERENTIAL ITEM FUNCTIONING USING SIMPLE VOLUME INDICES BASED ON MULTIDIMENTONAL ITEM RESPONSE THEORY**

*Heri Retnawati*

Fakultas Matematika dan Ilmu Pengetahuan Alam UNY  
retnawati\_heriuny@yahoo.co.id

### **Abstract**

The aim of the studi is to identify differential item functioning (DIF) based on item response theory using bidimensional model. This study uses explorative quantitative approach. Data used in this study is a mathematics test of national exam 2003 data which has been proved to contain 2-dimensions (Badrun Kartowagiran & Heri Retnawati, 2007). The simple volume indices are used to detect DIF which is the development of the simple area indices. The testee are grouped based on their gender, groups of male students and female students. The item parameter estimation was performed with BILOG and TESTFACT, and then to estimate volume and draw the surface characteristics curve was conduct with MAPLE. The results showed that of the 28 items analyzed, 2 items potentially contained uniform DIF and 26 items potentially contained a non-uniform DIF.

**Keywords:** *unidimensional IRT, multidimensional IRT, DIF, UPD, SPD*

## Pendahuluan

Idealnya, pelaksanaan tes berasaskan objektivitas, transparansi, akuntabilitas dan tidak diskriminatif. Jika pada suatu tes memuat suatu butir-butir soal yang memihak kelompok tertentu, maka tes tersebut dikatakan memuat bias atau mengandung keberfungsian butir diferensial (*differential item functioning, DIF*). Pada pendeteksian DIF, metode yang ada dan telah dikembangkan oleh ahli pengukuran/peneliti terdahulu masih berkisar pada analisis DIF dengan menggunakan teori respons butir unidimensi.

Unidimensi, artinya setiap butir tes hanya mengukur satu kemampuan. Asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung hanya satu komponen dominan yang mengukur prestasi suatu subjek. Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor administratif dalam tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Memperhatikan hal ini, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung hanya satu komponen dominan yang mengukur prestasi suatu subjek. Pada kenyataannya di lapangan, asumsi unidimensi sulit terpenuhi. Hal ini sesuai dengan pendapat bahwa kebanyakan tes pendidikan dan psikologi pada beberapa tingkat bersifat multidimensi (Bolt & Lall, 2003; Ackerman, et. al., 2003). Pada keadaan ini, analisis dengan pendekatan unidimensi sudah tidak sesuai lagi, dan akan mengakibatkan adanya kesalahan sistematis dan informasi yang diperoleh akan menyesatkan.

Dengan memperhatikan konsep dasar teori respons butir unidimensi dan multidimensi, dan prinsip metode dalam pendeteksian keberfungsian butir diferensial, dapat dikembangkan metode pendeteksian DIF berdasarkan teori respons butir multidimensi. Ada dua metode pendeteksian DIF yang telah dikembangkan, yakni dengan perbandingan *likelihood* (Retnawati, 2008), dan perbandingan probabilitas (Kartowagiran dan Retnawati, 2008). Pada studi ini

akan diidentifikasi DIF, berikut uji signifikansinya, diterapkan pada data real Ujian Nasional Pendidikan Matematika tahun 2003 yang telah diselidiki mengukur dua dimensi (Kartowagiran & Retnawati, 2007).

Pada teori respons butir unidimensi, hubungan antara tiga parameter butir yaitu indeks kesukaran butir soal, indeks daya beda butir, dan indeks tebakan semu (*pseudo-guessing*) dan satu kemampuan yang dinyatakan dalam persamaan peluang menjawab benar. Secara matematis, model logistik tiga parameter dapat dinyatakan sebagai berikut (Hambleton, dan Swaminathan, 1985 : 49; Hambleton, Swaminathan, dan Rogers, 1991: 17).

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

Keterangan:

- $\theta$  : tingkat kemampuan peserta tes
- $P_i(\theta)$  : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab butir  $i$  dengan benar
- $a_i$  : indeks daya beda dari butir ke- $i$
- $b_i$  : indeks kesukaran butir ke- $i$
- $c_i$  : indeks tebakan semu butir ke- $i$
- $e$  : bilangan natural yang nilainya mendekati 2,718
- $n$  : banyaknya item dalam tes
- $D$  : faktor penskalaan yang harganya 1,7.

Parameter  $b_i$  merupakan suatu titik pada skala kemampuan dalam kurva karakteristik butir ketika peluang menjawab benar peserta tes sebesar 50%. Parameter  $a_i$  merupakan indeks daya pembeda yang dimiliki butir ke- $i$ . Pada kurva karakteristik,  $a_i$  proporsional terhadap koefisien arah garis singgung (*slope*) pada titik  $\theta = b$ . Butir soal yang memiliki daya pembeda yang besar mempunyai kurva yang sangat menanjak, sedangkan butir soal yang mempunyai daya pembeda kecil mempunyai kurva yang sangat landai. Parameter ini menggambarkan probabilitas peserta dengan kemampuan rendah menjawab dengan benar pada suatu butir. Dengan adanya indeks tebakan semu

pada model logistik tiga parameter, memungkinkan subjek yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Nilai kemampuan peserta ( $\theta$ ) biasanya terletak di antara  $-3$  dan  $+3$ , sesuai dengan daerah asal distribusi normal. Gambar 1 menyajikan kurva karakteristik butir 1 ( $a=1, b=0,5, c=0$ ), butir 2 ( $a=0,5, b=0,5, c=0$ ) dan butir 3 ( $a=0,5, b=0,5, c=0,2$ ).

Model logistik 2 parameter dan model 1 parameter merupakan kasus dari model logistik 3 parameter. Ketika indeks *pseudo-guessing* sama dengan 0 ( $c=0$ ), model tersebut menjadi model 2 parameter. Demikian pula pada model 2 parameter, ketika indeks daya pembeda butir bernilai 1, maka model ini menjadi model logistik 1 parameter, atau lebih dikenal dengan nama model Rasch.

Pada teori respons butir multidimensi (*multidimensional item response theory, MIRT*) dikenal dua model, yakni *compensatory* dan *noncompensatory*. Menurut Ansley dan Forsyth (Spray, dkk., 1990), model *compensatory* membolehkan kemampuan tinggi pada salah satu dimensi memperoleh kompensasi pada kemampuan rendah pada dimensi lain dalam kaitannya dengan probabilitas menjawab benar. Sebaliknya, pada model *noncompensatory* tidak membolehkan kemampuan tinggi pada salah satu memperoleh kompensasi pada kemampuan rendah pada dimensi lainnya. Untuk model *compensatory* pada kasus butir dua dimensi, seorang peserta tes dengan kemampuan sangat rendah pada satu dimensi dan kemampuannya sangat tinggi pada dimensi lain dapat menjawab butir tes dengan benar.

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

2003) yang merupakan regresi logistik multivariat. Model MIRT logistik linear dapat ditulis sebagai :

$$P_i(\theta_j) = c_i + (1 - c_i) \frac{e^{\sum_{m=1}^k f_{ijm} + d_i}}{1 + e^{\sum_{m=1}^k f_{ijm} + d_i}} \dots(2)$$

Dengan  $f_{ijm} = \mathbf{a}_{im} \cdot \theta_{jm}$ ,  $c_i$  merupakan parameter *pseudo-guessing* butir ke- $i$ ,  $\mathbf{a}_{im}$  parameter diskriminasi untuk butir ke- $i$  pada dimensi ke- $m$ ,  $d_i$  parameter tingkat kesulitan butir ke- $i$ , dan  $\theta_{jm}$  merupakan elemen ke- $m$  dari vektor kemampuan orang ke  $j$  ( $\theta_j$ ). Senada dengan itu, Kirisci, Hsi dan Yu (2001) menuliskan persamaan (2) sebagai

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{(1 + e^{-1,7(a_i \theta_j - b_i)})} \dots\dots\dots(3)$$

yang memuat skala penyekalaan  $D=1,7$ .

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

$$P_i(\theta_j) = c_i + (1 - c_i) \prod_{m=1}^k \frac{e^{f_{ijm}}}{(1 + e^{f_{ijm}})} \dots\dots(4)$$

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

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

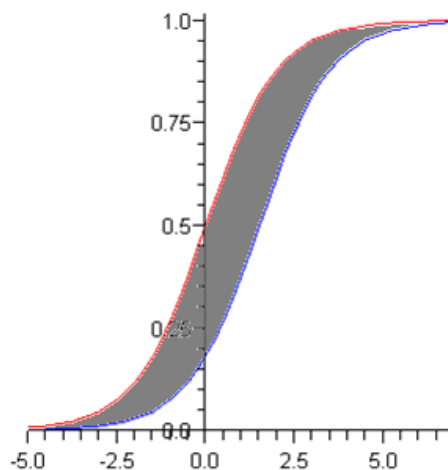
Parameter peserta tes pada model ini dinyatakan dengan elemen-elemen dari vektor  $\theta_j$ . Banyaknya elemen dari vektor ini

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

Diskriminasi butir pada teori respons butir multidimensi merupakan parameter untuk model yang dinyatakan dengan vektor **a** yang fungsinya mirip dengan parameter **a** pada teori respons butir unidimensi. Unsur-unsur vektor terkait dengan kemiringan dari permukaan respons pada arah yang bersesuaian dengan sumbu- $\theta$ . Kemiringan ini mengindikasikan sensitivitas butir terhadap kemampuan sepanjang sumbu- $\theta$ .

Konsep bias butir atau disebut juga keberfungsian butir pembeda (*differential item functioning*) didefinisikan sebagai perbedaan peluang menjawab benar antara dua kelompok yang dinamai grup Vokal dan grup Referensi (Angoff, 1993). Pada teori respons butir unidimensi, DIF dinyatakan sebagai perbedaan peluang menjawab benar suatu butir soal antara grup Vokal dan grup Referensi. Karena ukuran DIF dinyatakan dengan “seberapa besar perbedaan” antara kedua grup, pada kurva karakteristik ditandai dengan daerah yang diarsir pada Gambar 1. Daerah tersebut dinamai dengan daerah bertanda (SIGNED-AREA), yang ukuran luasnya dapat dihitung secara matematis dengan metode integrasi. Karena ukuran DIF terkait dengan ukuran luasan daerah sederhana, maka oleh Camilli dan Shepard (1994) metode ini dinamai dengan *Simple Area Indices*. Luas daerah yang diarsir disajikan pada persamaan (5).

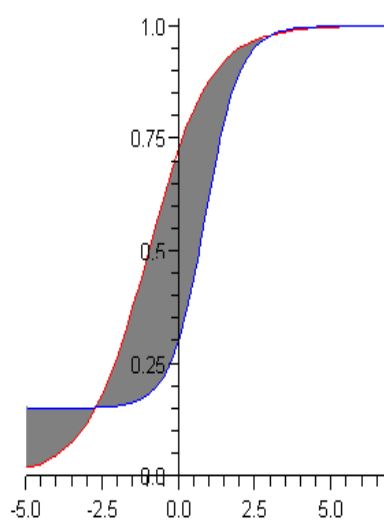
Pada Gambar 1, kurva karakteristik butir tidak saling memotong. Karena luas daerah merupakan integrasi dari peluang menjawab benar grup referensi dikurangi dengan grup vokal, maka jika bernilai positif, butir soal menguntungkan kelompok referensi, Sebaliknya, jika bernilai negatif, butir soal menguntungkan grup vokal.



Gambar 1. Kurva Karakteristik Butir 2 Kelompok yang Tidak Saling Memotong

$$SIGNED - AREA = \int [P_R(\theta) - P_F(\theta)] d\theta \dots\dots(5)$$

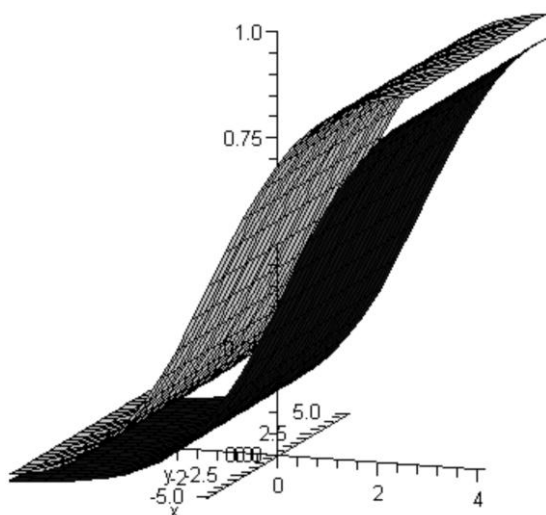
Dalam analisis DIF suatu butir, bisa jadi kurva karakteristik butir dari kedua grup saling berpotongan. Jika hal ini terjadi, ukuran DIF yang positif dan yang negatif akan saling meniadakan, seperti yang digambarkan pada Gambar 2. Pada kasus ini, ukuran luasan dapat dihitung dengan UNSIGNED-AREA yang merupakan integral dari kuadrat selisih antara peluang menjawab benar grup referensi dengan grup vokal.



Gambar 2. Kurva Karakteristik Butir dari Dua Kelompok yang Saling Memotong

$$UNSIGNED - AREA = \sqrt{\int [P_R(\theta) - P_F(\theta)]^2 d\theta} \dots (6)$$

Dengan menggunakan konsep perbedaan peluang menjawab benar antara grup referensi dengan grup vokal, konsep ini dapat digunakan pada fungsi peluang menjawab benar yang multivariat. Konsep luasan sebagai perbedaan peluang bergeser menjadi perbedaan volume, sehingga *Simple Area Indices* pada IRT unidimensi dikembangkan menjadi *Simple Volume Indices* pada teori respons butir multidimensi. Pada butir yang mengukur dua dimensi kemampuan, misalnya  $\theta_1$  dan  $\theta_2$ , permukaan karakteristik digambarkan pada Gambar 3.



Gambar 3. Permukaan Karakteristik Butir dari Dua Kelompok yang Tidak Saling Memotong

Daerah di antara kedua permukaan karakteristik tersebut dinamai dengan volume bertanda (*SIGNED-VOLUME*), yang ukuran volumenya dapat dihitung secara matematis dengan metode integrasi ganda.

$$SIGNED - VOLUME =$$

$$\int \int [P_R(\theta_1, \theta_2) - P_F(\theta_1, \theta_2)] d\theta_1 d\theta_2 \dots (7)$$

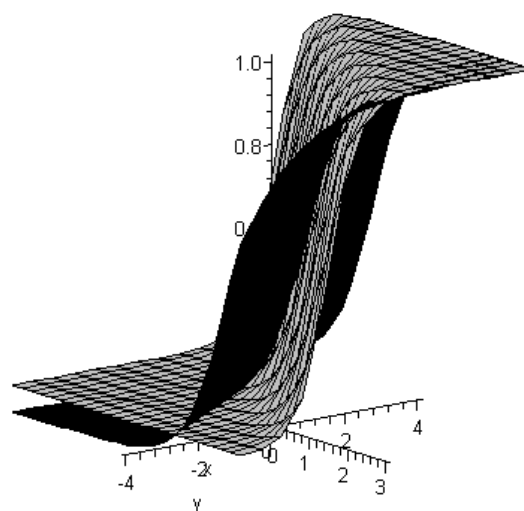
Pada Gambar 3 di atas, permukaan karakteristik butir tidak saling memotong. Karena volume tersebut merupakan integrasi dari peluang menjawab benar grup referensi dikurangi dengan grup vokal,

maka jika bernilai positif, butir soal menguntungkan kelompok referensi, namun sebaliknya, jika bernilai negatif, butir soal menguntungkan grup vokal.

Dalam analisis DIF suatu butir, bisa jadi permukaan karakteristik butir dari kedua grup saling berpotongan. Jika hal ini terjadi, ukuran DIF yang positif dan yang negatif juga akan saling meniadakan, seperti yang digambarkan pada Gambar 4. Pada kasus ini, ukuran luasan dapat dihitung dengan *UNSIGNED-VOLUME* yang merupakan integral dari kuadrat selisih antara peluang menjawab benar grup referensi dengan grup vokal.

$$UNSIGNED - VOLUME =$$

$$\sqrt{\int \int [P_R(\theta_1, \theta_2) - P_F(\theta_1, \theta_2)]^2 d\theta_1 d\theta_2} \dots (8)$$



Gambar 4. Permukaan Karakteristik Butir dari Dua Kelompok yang Saling Memotong

Kemiripan sifat ini dapat digunakan untuk butir yang mengukur lebih dari 2 dimensi, namun sudah tidak dapat digambarkan lagi. Untuk butir yang mengukur kemampuan  $\theta_1, \theta_2, \dots, \theta_k$ , *SIGNED-VOLUME* dan *UNSIGNED-VOLUME* dapat ditentukan dengan

$$SIGNED - VOLUME =$$

$$\int \int \dots \int [P_R(\theta_1, \theta_2, \dots, \theta_k) - P_F(\theta_1, \theta_2, \dots, \theta_k)] d\theta_1 d\theta_2 \dots d\theta_k \dots (9)$$

untuk permukaan karakteristik yang tidak saling berpotongan dan

UNSIGNED – VOLUME =

$$\sqrt{\int \int \dots \int [P_R(\theta_1, \theta_2, \dots, \theta_k) - P_F(\theta_1, \theta_2, \dots, \theta_k)]^2 d\theta_1 d\theta_2 \dots d\theta_k} \dots \dots \dots (10)$$

untuk permukaan karakteristik yang saling berpotongan.

Untuk mengetahui sigifikansi ukuran DIF ini, dapat digunakan perbandingan likelihood antara dua model. Metode ini merupakan metode untuk mengetahui signifikansi DIF dengan menggunakan Teori Respons Butir, dan oleh Camilli dan Shepard (1994 : 76-97) diberi nama dengan mengetahui perbandingan model dan menurut Thissen et. al. (1993: 72) diberi nama dengan Metode IRT-LR. Langkah-langkah untuk mendeteksi DIF sebagai berikut.

Misalkan  $L^*$  merupakan nilai fungsi likelihood  $L$ . Ada dua model yang akan diperbandingkan, model C, yaitu model kompak (*compact*) dan model A, yaitu model yang ditingkatkan (*augmented*). Model C merupakan model yang lebih sederhana. Kemudian dirumuskan hipotesis :

$$H_o : \Gamma = Set_C \text{ (Set}_C \text{ memuat N parameter)} \dots \dots \dots (11)$$

$$H_a : \Gamma = Set_A \text{ (Set}_A \text{ memuat N+M parameter)} \dots (12)$$

$\Gamma$  dianggap memiliki set parameter yang benar. Model C memiliki M parameter lebih sedikit dibandingkan dengan model A. Perbandingan *likelihood (Likelihood Ratio, LR)* untuk dua model dinyatakan dengan persamaan:

$$LR = \frac{L^*_{(C)}}{L^*_{(A)}} \dots \dots \dots (13)$$

dengan:

$L^*_{(C)}$  : nilai fungsi *likelihood* model C

$L^*_{(A)}$  : nilai fungsi *likelihood* model A.

Kemudian ditransformasikan dengan logaritma natural:

$$\begin{aligned} x^2_{(M)} &= -2 \ln(LR) \\ &= [-2 \ln L^*_{(C)}] - [-2 \ln L^*_{(A)}] \dots (14) \end{aligned}$$

dengan:

$L^*_{(C)}$  : nilai fungsi likelihood model C

$L^*_{(A)}$  : nilai fungsi likelihood model A.

Agar lebih mudah,  $G(C) = [-2 \ln L^*_{(C)}]$  dan  $G(A) = [-2 \ln L^*_{(A)}]$ , sehingga rasio/perbandingan logaritma likelihood menjadi

$$x^2_{(M)} = -2 \ln(LR) = G(C) - G(A) \dots (15)$$

Seperti yang dinyatakan oleh Camilli dan Shepard (1994 : 76), *if we take the natural log transformation of the LR, we end up with a test statistic that is approximately distributed as a chi-square with M degrees of freedom.*

Langkah-langkah untuk mendeteksi DIF, dapat diuraikan sebagai berikut. Pertama, menaksir parameter butir dengan teori respons butir model 3P, memperoleh statistik kecocokan yang mendekati distribusi  $\chi^2$  transformasi kebolehjadian  $G(C)$  pada tes yang terdiri dari K butir. Kedua, ditentukan salah satu butir tes, misalnya butir ke-k, untuk dievaluasi. Ketiga, butir tes tersebut dibuat seolah-olah menjadi dua butir. Butir pertama berisi jawaban dari salah satu kelompok, misalnya kelompok referensi, yang tidak direspons oleh kelompok vokal. Butir kedua berisi jawaban dari kelompok vokal yang tidak direspons oleh kelompok referensi. Keempat, menaksir kembali parameter, dan diperoleh harga  $\chi^2$  transformasi kebolehjadian  $G(A)$  untuk tes yang terdiri dari K+1 butir. Selanjutnya, dapat ditentukan nilai  $\chi^2_{(M)}$ , yang dapat digunakan untuk mengetahui signifikansi keberadaan DIF pada suatu butir soal.

### Metode Penelitian

Studi eksplorasi ini menggunakan pendekatan kuantitatif, dalam rangka mendeteksi muatan keberfungsian butir pembeda (*differential item functioning, DIF*) pada perangkat tes. Data yang digunakan pada penelitian ini merupakan data Ujian Nasional

mata pelajaran matematika tahun 2003 yang sudah dibuktikan memuat 2 dimensi (Kartowagiran & Retnawati, 2007). Selanjutnya, estimasi parameter untuk mengestimasi  $c$  (*pseudo guessing*) dilakukan dengan BILOG (Mislevy & Bock, 1990), analisis dengan teori respons butir multidimensi dengan TESTFACT, analisis perbandingan *Likelihood* dengan TESTFACT (Wilson, D., Wood, R. & Gibbons, R., 1984), dan analisis perbandingan peluang dengan MAPLE. Hasil analisis dilakukan dengan menggunakan sintaks program yang telah disusun oleh Kartowagiran & Retnawati (2009).

**Hasil dan Pembahasan**

Dari 40 butir tes Ujian Nasional Mata Pelajaran Matematika, ternyata hanya 28 butir tes yang cocok dengan model logistik 3 parameter. Selanjutnya, ke-28 butir tes inilah yang akan dideteksi muatan keberfungsian butir pembedanya. Selanjutnya,

hanya butir-butir yang cocok dengan model saja yang selanjutnya dianalisis muatan DIF-nya.

Selanjutnya, untuk mengestimasi parameter tiap kelompok, terlebih dahulu diestimasi parameter  $c$  dengan BILOG, yang hasilnya dientrikan pada TESTFACT. Selanjutnya dilakukan estimasi parameter terpisah untuk dapat mengestimasi perbedaan luas di bawah kurva antara kelompok laki-laki dan kelompok perempuan, dilakukan estimasi terpisah dari data respons peserta di kedua kelompok. Hasil estimasi disajikan pada Tabel 1 untuk kelompok laki-laki dan perempuan.

Setelah dilakukan analisis yang unidimensi, dilakukan analisis bidimensi untuk data respons peserta UN mata pelajaran Matematika 2003. Analisis dilakukan dengan bantuan *software* TESTFACT dari Ssi. Hasil analisis disajikan pada Tabel 1.

Tabel 1. Parameter Butir UN 2003 Mapel Matematika 2 Dimensi

Butir	c	b	a1	a2	Butir	c	b	a1	a2
2	0,155	-0,243	1,066	0,129	23	0,208	-0,162	0,983	0,383
3	0,207	0,518	0,824	0,146	24	0,153	-0,319	0,944	0,041
4	0,1	-1,487	0,958	0,084	25	0,272	0,482	1,436	0,096
5	0,27	0,348	1,204	-0,048	26	0,351	0,413	1,27	0,163
8	0,054	-0,322	1,448	-0,163	27	0,325	-1,389	1,2	0,441
9	0,158	-0,191	1,117	-0,125	29	0,325	0,093	1,183	0,017
10	0,273	0,747	1,468	-0,144	30	0,144	-0,401	0,861	-0,014
11	0,149	-0,411	0,777	0,029	31	0,214	-1,062	1,063	0,33
12	0,294	0,484	0,779	0,055	32	0,237	1,133	0,713	0,23
18	0,202	-1,716	1,179	0,458	33	0,458	-2,632	1,502	0,964
19	0,28	-1,943	1,547	-0,166	36	0,315	-0,594	0,811	0,363
20	0,231	-1,805	1,396	-0,109	37	0,257	-1,184	0,905	0,54
21	0,202	0,918	1,795	0,525	39	0,115	-0,526	0,875	0,194
22	0,177	0,202	1,533	0,664	40	0,226	0,179	1,093	0,472

Seperti halnya pada analisis DIF dengan teori respons butir unidimensi, pada analisis DIF dengan pendekatan teori respons butir multidimensi terlebih dahulu

dilakukan dilakuan analisis terpisah kelompok wanita dan pria. Hasilnya disajikan pada Tabel 2.

Tabel 2. Parameter Berdasarkan Teori Respons Butir Multidimensi pada Kelompok Pria dan Wanita

No. Butir	Pria				Wanita			
	c	b	a1	a2	c	b	a1	a2
2	0,155	-0,431	1,132	0,148	0,155	-0,032	0,944	0,219
3	0,207	0,446	0,829	0,22	0,207	0,602	0,801	0,198
4	0,1	-1,598	0,916	0,047	0,1	-1,332	0,945	0,169
5	0,27	0,305	1,24	0,127	0,27	0,402	1,199	-0,243
8	0,054	-0,597	1,511	-0,11	0,054	-0,033	1,396	-0,23
9	0,158	-0,191	1,026	-0,062	0,158	-0,225	1,427	-0,43
10	0,273	0,548	1,441	0,011	0,273	0,911	1,361	-0,031
11	0,149	-0,471	0,879	-0,182	0,149	-0,386	0,7	0,187
12	0,294	0,288	0,9	0,083	0,294	0,695	0,638	0,097
18	0,202	-1,886	1,202	0,461	0,202	-1,551	1,19	0,333
19	0,28	-2,814	2,255	-0,809	0,28	-1,792	1,327	0,282
20	0,231	-1,742	1,357	-0,131	0,231	-1,774	1,3	-0,05
21	0,202	1,11	2,337	1,115	0,202	0,905	1,63	0,118
22	0,177	0,25	1,777	1,017	0,177	0,18	1,487	0,453
23	0,208	-0,198	0,931	0,264	0,208	-0,11	1,018	0,343
24	0,153	-0,446	1,011	0,123	0,153	-0,193	0,895	-0,091
25	0,272	0,465	1,709	0,145	0,272	0,496	1,196	0,175
26	0,351	0,411	1,256	0,119	0,351	0,411	1,359	0,066
27	0,325	-1,746	1,392	0,321	0,325	-0,835	0,802	0,442
29	0,325	0,047	1,361	-0,081	0,325	0,129	1,009	0,076
30	0,144	-0,545	0,851	-0,056	0,144	-0,237	0,844	0,07
31	0,214	-1,184	1,189	0,297	0,214	-0,949	0,971	0,195
32	0,237	1,103	0,659	0,374	0,237	1,221	0,808	0,07
33	0,458	-2,34	1,383	0,505	0,458	-2,387	1,376	0,865
36	0,315	-0,587	0,786	0,347	0,315	-0,595	0,863	0,201
37	0,257	-1,062	0,735	0,308	0,257	-1,155	1,016	0,485
39	0,115	-0,615	0,869	0,01	0,115	-0,435	0,891	0,355
40	0,226	0,197	1,116	0,509	0,226	0,158	1,077	0,358

Berdasarkan parameter hasil estimasi tersebut, selanjutnya dilakukan identifikasi muatan DIF dengan pendekatan teori respons butir multidimensi, sekaligus analisis signifikansinya.

Seperti halnya pada deteksi muatan DIF pada butir dengan teori respons butir unidimensi, untuk mengetahui muatan DIF dengan teori respons butir multidimensi dapat diketahui dengan mengestimasi indeks DIF dan menguji signifikansi dari indeks DIF. Indeks DIF diketahui dengan menghitung perbedaan volume dan perbedaan peluang dari sejumlah kemampuan hasil estimasi peserta tes. Menguji signifikansi

dilakukan dengan mengetahui perbedaan probabilitas dengan menggunakan perbandingan *likelihood*.

#### Mengestimasi Indeks DIF

Untuk dapat menentukan hubungan antara permukaan karakteristik butir pada kelompok laki-laki dan kelompok perempuan, terlebih dahulu digambarkan kedua kurva tersebut pada ruang XYZ yang sama. Menggambar permukaan karakteristik dilakukan dengan bantuan MAPLE 10.0. Indeks perbedaan volume di bawah kurva juga dihitung dengan menghitung nilai integral dengan pendekatan yang juga dilakukan



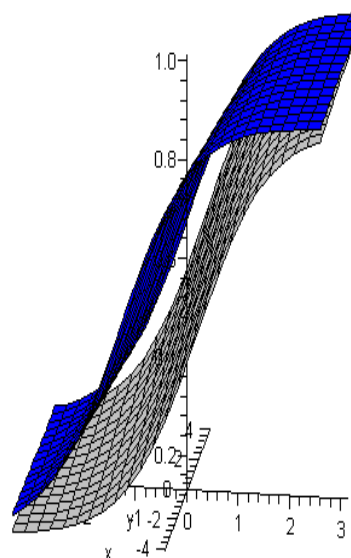
dengan MAPLE 10.0. Hasilnya disajikan pada Tabel 3.

Tabel 3. Hasil Estimasi Indeks dengan Perbedaan Volume di Bawah Kurva

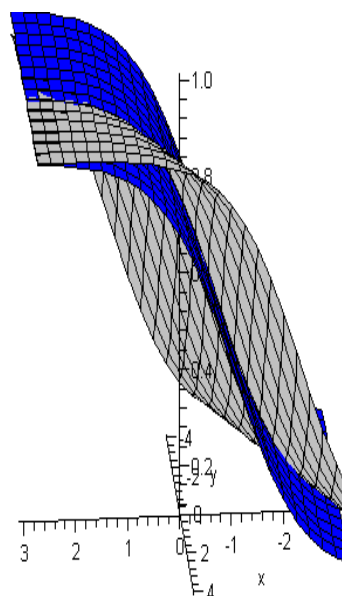
Butir	Hubungan Kedua		UPD	SPD
	Kurva			
2	Tidak Berpotongan		2,376	0,320
3	Berpotongan		1,334	0,478
4	Berpotongan		2,381	0,341
5	Berpotongan		0,468	0,713
8	Berpotongan		3,051	0,699
9	Berpotongan		0,245	0,624
10	Berpotongan		1,561	0,104
11	Berpotongan		-0,052	1,426
12	Berpotongan		4,156	0,567
18	Berpotongan		1,646	0,193
19	Berpotongan		-0,488	2,634
20	Berpotongan		-0,497	0,0451
21	Berpotongan		0,592	2,223
22	Berpotongan		-0,180	0,968
23	Berpotongan		0,658	0,039
24	Berpotongan		1,539	0,483
25	Berpotongan		0,623	0,054
26	Berpotongan		-0,076	0,014
27	Berpotongan		1,333	0,686
29	Berpotongan		0,497	0,161
30	Tidak Berpotongan		6,286	1,327
31	Berpotongan		0,144	0,040
32	Berpotongan		-0,396	1,070
33	Berpotongan		-0,080	0,355
36	Berpotongan		0,262	0,203
37	Berpotongan		1,618	0,184
39	Berpotongan		1,564	1,227
40	Berpotongan		-0,195	0,118

Gambar permukaan karakteristik kelompok laki-laki dan kelompok perempuan yang tidak berpotongan mengindikasikan bahwa butir tersebut menguntungkan suatu kelompok di semua daerah kemampuan (memuat DIF uniform). Sebaliknya, jika permukaan kedua permukaan berpotongan, mengindikasikan butir menguntungkan kelompok laki-laki di suatu wilayah kemampuan, juga menguntungkan kelompok lain di wilayah kemampuan lainnya (memuat DIF

nonuniform). Sebagai contoh, butir soal nomor 30 memuat DIF uniform dan butir nomor 39 memuat DIF nonuniform yang disajikan pada Gambar 5 dan Gambar 6.



Gambar 5. Permukaan Karakteristik Butir Kelompok Laki-laki (permukaan gelap) dan Kelompok Perempuan (permukaan abu-abu) pada Butir Soal Nomor 30 (DIF uniform)



Gambar 6. Permukaan Karakteristik Butir Kelompok Laki-laki (permukaan gelap) dan Kelompok Perempuan (permukaan abu-abu) pada Butir Soal Nomor 39 (DIF nonuniform)

Menguji Signifikansi

Menguji signifikansi indeks DIF dilakukan dengan perbandingan *likelihood*. Untuk dapat membandingkan *likelihood* antara model yang ditingkatkan (G(A)), dan pada model kompak (G(C)), terlebih dahulu di-

estimasi *likelihood*-nya. Estimasi *likelihood* dilakukan dengan bantuan *software* TEST-FACT dari Ssi. Hasil G(C)-G(A) selanjutnya dibandingkan dengan Khi-kuadrat tabel. Hasil selengkapnya disajikan pada Tabel 16.

Tabel 4. Signifikansi DIF dengan Perbandingan *Likelihood* (2 dimensi)

Butir	G(A)	G(C)	G(C)-G(A)	dk	Kesimpulan
2	8.68E+04	9.03E+04	3530	25	Signifikan Memuat DIF
3	8.68E+04	9.09E+04	4108.29	25	Signifikan Memuat DIF
4	8.68E+04	8.83E+04	1530.963	25	Signifikan Memuat DIF
5	8.68E+04	9.17E+04	4935.406	25	Signifikan Memuat DIF
8	8.68E+04	8.97E+04	2909.738	25	Signifikan Memuat DIF
9	8.68E+04	9.04E+04	3652.778	25	Signifikan Memuat DIF
10	8.68E+04	9.18E+04	5043.338	25	Signifikan Memuat DIF
11	8.68E+04	8.96E+04	2848.448	25	Signifikan Memuat DIF
12	8.68E+04	9.18E+04	5010.209	25	Signifikan Memuat DIF
18	8.68E+04	8.90E+04	2.20E+03	25	Signifikan Memuat DIF
19	8.68E+04	Tidak Konvergen			
20	8.68E+04	8.96E+04	2.77E+03	25	Signifikan Memuat DIF
21	8.68E+04	9.17E+04	4.92E+03	25	Signifikan Memuat DIF
22	8.68E+04	9.04E+04	3.59E+03	25	Signifikan Memuat DIF
23	8.68E+04	9.02E+04	3.37E+03	25	Signifikan Memuat DIF
24	8.68E+04	8.96E+04	2.81E+03	25	Signifikan Memuat DIF
25	8.68E+04	9.13E+04	4.47E+03	25	Signifikan Memuat DIF
26	8.68E+04	9.16E+04	4.77E+03	25	Signifikan Memuat DIF
27	8.68E+04	9.02E+04	3.39E+03	25	Signifikan Memuat DIF
29	8.68E+04	9.13E+04	4.50E+03	25	Signifikan Memuat DIF
30	8.68E+04	8.94E+04	2.60E+03	25	Signifikan Memuat DIF
31	8.68E+04	8.92E+04	2.41E+03	25	Signifikan Memuat DIF
32	8.68E+04	9.17E+04	4.93E+03	25	Signifikan Memuat DIF
33	8.68E+04	Tidak Konvergen			
36	8.68E+04	9.00E+04	3.20E+03	25	Signifikan Memuat DIF
37	8.68E+04	8.95E+04	2.67E+03	25	Signifikan Memuat DIF
39	8.68E+04	8.98E+04	2994.733	25	Signifikan Memuat DIF
40	8.68E+04	9.06E+04	3766.21	25	Signifikan Memuat DIF

Mencermati hasil tersebut, dapat diperoleh bahwa 26 butir yang signifikan memuat DIF, 2 butir yang tidak konvergen pada analisisnya, dan tidak ada butir yang netral dari 28 butir yang dianalisis.

Setelah diterapkan pada data, indeks DIF yang ditentukan dengan UPD maupun SPD tidak memberikan hasil yang terlalu berbeda. Hal ini disebabkan karena indeks UPD dan SPD dengan yang dihitung dengan perbedaan peluang merupakan bagian dari volume yang dihitung, semakin banyak titik kemampuan yang dibangkitkan, akan semakin mendekati volume.

Indeks UPD dan SPD yang dihitung dengan perbedaan peluang, signifikansinya tergantung variabilitas dalam kelompok dan sangat tergantung pada varians interkelompok. Hal ini menyebabkan hasil yang berbeda jika pengambilan anggota subkelompok acak, dan estimasi UPD dan SPD pada subkelompok tidak stabil.

Berdasarkan hasil analisis, diperoleh bahwa dengan metode *likelihood* diperoleh hasil yang lebih stabil. Pada analisis DIF khususnya pada model 2 dimensi, program sangat peka terhadap sifat data. Misalnya pada data yang kurang baik, data yang matriks korelasi antarvariabel merupakan matriks singular, analisis yang dijalankan tidak konvergen. Akibat ketidakkonvergenan ini yakni tidak bisa teruji signifikansi DIFnya, karena nilai maksimum *likelihood*-nya tidak bisa diperoleh.

Pada identifikasi DIF dengan model 2 dimensi hampir semua signifikan memuat DIF. Hal tersebut disebabkan karena model 2 dimensi lebih teliti dalam membandingkan probabilitas kemampuan. Model ini sangat sensitif karena pada model dua dimensi, perbedaan yang kecil akan menjadi kelihatan pada gambar permukaan. Uji- $\chi^2$  yang digunakan pada penelitian ini juga sangat peka dengan ukuran sampel. Perbedaan sedikit pada peluang, dengan ukuran sampel yang besar perubahan ini menjadi signifikan. Hal inilah yang mengakibatkan model analisis dengan perbandingan *likelihood* sangat peka untuk mengidentifikasi DIF.

## Simpulan dan Rekomendasi

Berdasarkan hasil analisis, dapat disimpulkan bahwa dari 28 butir yang dianalisis, semuanya berpotensi memuat DIF berdasarkan pendekatan teori respons butir bidimensi. Sebanyak 2 butir berpotensi memuat DIF uniform dan 26 butir berpotensi memuat DIF nonuniform. Dengan menggunakan perbandingan *likelihood*, 28 butir tersebut signifikan memuat DIF yang menguntungkan kelompok siswa tertentu jika dikategorikan berdasarkan jenis kelaminnya, laki-laki dan perempuan.

Mencermati hasil penelitian ini, dapat direkomendasikan

1. Penelitian ini baru mengetahui signifikansi DIF dengan mengaplikasikan metode pada data real, belum dengan simulasi untuk menyelidiki sifat-sifat DIF, sehingga terkait dengan penelitian di masa mendatang, dapat dilakukan penelitian terkait perbandingan sensitivitas metode mendeteksi DIF dengan simulasi yang melibatkan jumlah butir, banyaknya peserta, banyaknya dimensi.
2. Data yang dipergunakan pada penelitian ini hanya mengukur 2 dimensi, penelitian yang menggunakan data lebih dari 2 dimensi juga perlu dilakukan.
3. Penelitian mendeteksi DIF sulit dilakukan, terkait dengan banyaknya tahap yang harus dilalui, sehingga perlu dikembangkan *batch-file* untuk memerintah komputer agar beberapa program dapat berjalan otomatis dan secara otomatis pula membaca hasil suatu *output* untuk menjadi input program lainnya.

## Daftar Pustaka

- Ackerman, T.A., et. al. (2003). Using multi-dimensional item response theory to evaluate educational and psychological tests. *Educational Measurement*, 22, 37-53.
- Angoff, W.H. 1993. Perspectives on differential item functioning methodology. Dalam P.W. Holland dan H. Wainer

- (Eds.), *Differential item functioning*. Hillsdale, NJ: Erlbaum, Pp. 3 – 23.
- Bolt, D.M. & Lall, V.M. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte-Carlo. *Applied Psychological Measurement*, 27, 395-414.
- Camilli, G. & Shepard, L.A. 1994. *Methods for identifying biased test items, Vol.4*. London: Sage Publications, Inc.
- Hambleton, R.K. & Rogers, H.J. 1995. Developing an item bias review form. From <http://www.ericcae.net/ft/ta/ta/biaspub2.htm> March 10, 2007.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamentals of item response theory*. London: Sage Publications, Inc.
- Hambleton, R.K. & Swaminathan. 1985. *Item response theory*. Boston, MA: Kluwer Nijhoff, Publisher.
- Kartowagiran, B & Retnawati, H. (2007). Validasi Dimensionalitas Perangkat Tes UN 2003-2006. *Laporan Penelitian*. Lembaga Penelitian Universitas Negeri Yogyakarta.
- Kartowagiran, B & Retnawati, H. (2008). The Probability Difference Indices and Empirical Sampling Distribution for DIF Indices for Identifying Item Bias In Multidimensional Item Response Theory. *Paper on International Conference on Mathematics 3th on July 18<sup>th</sup> 2008*.
- Kartowagiran, B & Retnawati, H. (2009). Pengembangan Model Identifikasi Differential Item Functioning dengan Teori Respons Butir Multidimensi. *Laporan Penelitian*. Lembaga Penelitian Universitas Negeri Yogyakarta.
- Retnawati, H. (2008). Identifying Item Bias Using the Simple Volume Indices and Multidimensional Item Response Theory Likelihood Ratio (IRT-LR) Test. *Paper on International Conference on Mathematics 3th on July 18<sup>th</sup> 2008*.
- Mislevy, R.J. & Bock, R.D. (1990). *BILOG 3: Item analysis & test scoring with binary logistic models*. Mooreville: Scientific Software Inc.
- Reckase, M.D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W.J. Linden & R.K. Hambleton (Eds), *Handbook of modern item response theory* (pp. 271-286). New York: Springer.
- Spray, J.A., et al. (1990). Comparison of two logistic multidimensional item response theory models. *ACT Research Report Series*. United States Government.
- Thissen, D. et al. (1993). Detection of differential item functioning using the parameters of item response model. Dalam P.W. Holland & H. Wainer (Ed). *Differential Item Functioning*. Englewood Clifp, NJ : Lawrence Erlbaum. p.67-86.
- Wilson, D., Wood, R. & Gibbons, R. (1984). *TESTFACT: Test scoring and fullinformation item factor analysis*. [Computer program]. Mooresville, IL: Ssi.