

**Mengidentifikasi Bias Butir Perangkat Tes Matematika UN 2003 dengan
Menggunakan Indeks Volume Sederhana dan Perbandingan Likelihood
pada Teori Respons Butir Multidimensi**

Heri Retnawati

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi bias butir (differential item functioning) pada perangkat tes Matematika UN 2003 dengan menggunakan perbedaan peluang antara siswa kelompok pria dan kelompok wanita. Ukuran DIF dapat diestimasi dengan perbedaan probabilitas untuk menjawab benar pada kedua grup, yang disebut grup vocal dan grup referensi. Pada data multidimensi, perbedaan peluang dapat diketahui dengan menggunakan indeks volume sederhana. Signifikansi dari ukuran DIF antara dua kelompok dapat diuji dengan menggunakan uji perbandingan likelihood antara model kompak yang memuat butir yang diteliti dan model keseluruhan yang memuat keseluruhan butir. Hasil penelitian menunjukkan bahwa

Kata Kunci : MIRT, DIF, Indeks volume sederhana, Uji perbandingan likelihood

Pendahuluan

Tes merupakan salah satu alat pengukuran yang paling sering digunakan pada bidang pendidikan dan psikometri. Pada pelaksanaannya, tes seharusnya 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 subyek. 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 subyek. 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 dan Lall, 2003; Ackerman, dkk., 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, telah dikembangkan metode pendeteksian DIF berdasarkan teori respons butir multidimensi (Heri Retnawati, 2008; Badrun Kartowagiran dan Heri Retnawati, 2008). Pada tulisan ini akan dibahas metode pendeteksian DIF dengan menggunakan prinsip Simple Area Indices pada IRT unidimensi yang dikembangkan menjadi Simple Volume Indices pada IRT Multidimensi, dan perbandingan likelihood untuk mengetahui signifikansinya, dan penerapannya untuk mengidentifikasi DIF pada perangkat matematika ujian nasional 2003.

Kerangka Teoretis

Pada teori respons butir unidimensi, hubungan antara tiga parameter butir yaitu indeks kesukaran butir soal, indeks daya beda butir, dan indeks tebakan semu (*pseudoguessing*) 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).

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

Keterangan :

θ : tingkat kemampuan peserta tes

$P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ 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 (θ) 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 :

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

Dengan $f_{ijm} = \mathbf{a}_{jm} \cdot \theta_{im}$, c_i merupakan parameter *pseudo-guessing* butir ke-i, \mathbf{a}_{jm} parameter diskriminasi untuk butir ke-i pada dimensi ke-m, d_i parameter tingkat kesulitan butir ke-i, dan θ_{jm} merupakan elemen ke-m dari vektor kemampuan orang ke j (θ_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 :

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

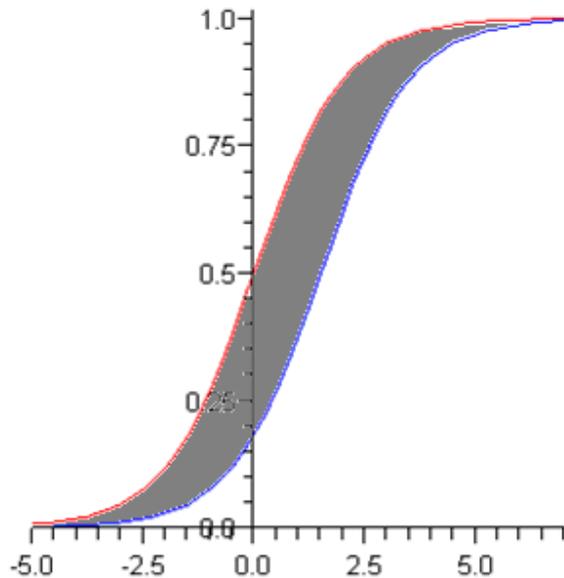
Dengan $f_{ijm} = [\mathbf{a}_{im} (\boldsymbol{\theta}_{jm} - b_{im})]$ dengan b_{im} merupakan parameter tingkat kesulitan butir ke- i pada dimensi ke- m . Terkait dengan bentuknya yang merupakan hasil perkalian, model ini sering pula dinamai dengan model multiplikatif.

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 $\boldsymbol{\theta}_j$. 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 \mathbf{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- θ . Kemiringan ini mengindikasikan sensitivitas butir terhadap kemampuan sepanjang sumbu- θ .

Konsep bias butir atau disebut juga keberfungsian butir pembeda (*differential item functioning*) didefinisikan sebagai perbedaan peluang menjawab benar antara dua kelompok yang dinamai grup Fokal dan grup Referensi. Pada teori respons butir unidimensi, DIF dinyatakan sebagai perbedaan peluang menjawab benar suatu butir soal antara grup Fokal dan grup Referensi. Karena ukuran DIF dinyatakan dengan “seberapa besar perbedaan” antara kedua grup, pada kurva karakteristik ditandai dengan daerah yang diarsir pada gambar 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*.



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

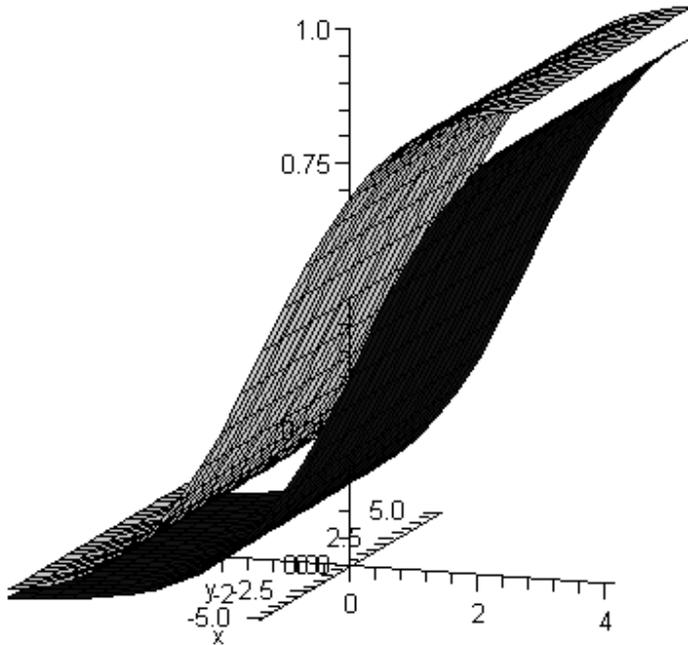
Pada gambar di atas, kurva karakteristik butir tidak saling memotong. Karena luas daerah merupakan integrasi dari peluang menjawab benar grup referensi dikurangi dengan grup fokal, maka jika bernilai positif, butir soal menguntungkan kelompok referensi, Sebaliknya, jika bernilai negatif, butir soal menguntungkan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

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 Pada kasus ini, ukuran luasan dapat dihitung dengan UNSIGNED-AREA yang merupakan integral dari kuadrat selisih antara peluang menjawab benar grup referensi dengan grup fokal.

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

Dengan menggunakan konsep perbedaan peluang menjawab benar antara grup referensi dengan grup fokal, 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 θ_1 dan θ_2 , permukaan karakteristik digambarkan pada Gambar



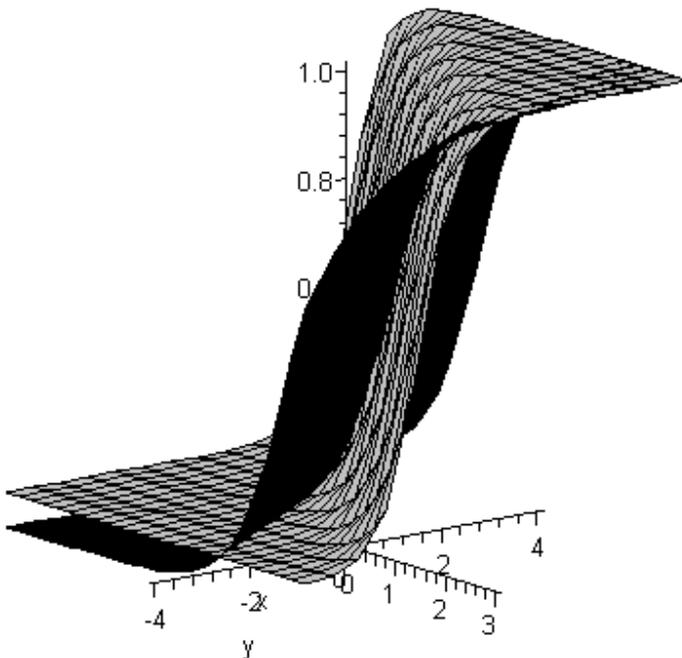
Daerah di antara kedua permukaan karakteristik tersebut dinamai dengan volume bertanda (SIGNED-VOLUME), yang ukuran volumenya dapat dihitung secara matematis dengan metode integrasi ganda.

$$\text{SIGNED-VOLUME} = \iint [P_R(\theta_1, \theta_2) - P_F(\theta_1, \theta_2)] d\theta_1 d\theta_2 \dots\dots\dots(7)$$

Pada gambar di atas, permukaan karakteristik butir tidak saling memotong. Karena volume tersebut merupakan integrasi dari peluang menjawab benar grup referensi dikurangi dengan grup fokal, maka jika bernilai positif, butir soal menguntungkan kelompok referensi, namun sebaliknya, jika bernilai negatif, butir soal menguntungkan grup fokal.

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 Pada kasus ini, ukuran luasan dapat dihitung dengan UNSIGNED-VOLUME yang merupakan integral dari kuadrat selisih antara peluang menjawab benar grup referensi dengan grup fokal.

$$\text{UNSIGNED-VOLUME} = \sqrt{\iint [P_R(\theta_1, \theta_2) - P_F(\theta_1, \theta_2)]^2 d\theta_1 d\theta_2} \dots\dots\dots(8)$$



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 SIGNED-VOLUME dapat ditentukan dengan

$$\text{SIGNED-VOLUME} = \iiint \dots \int [P_R(\theta_1, \theta_2, \dots, \theta_k) - P_F(\theta_1, \theta_2, \dots, \theta_k)] d\theta_1 d\theta_2 \dots d\theta_k \dots\dots\dots(9)$$

untuk permukaan karakteristik yang tidak saling berpotongan dan

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

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 :

$$H_0 : \Gamma = \text{Set}_C \text{ (Set}_C \text{ memuat N parameter)} \dots \dots \dots (11)$$

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

Γ dianggap memiliki set parameter yang benar. Model C memiliki M parameter lebih sedikit dibandingkan dengan model A. Perbandingan likekihood (*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 :

$$\chi^2_{(M)} = -2 \ln(LR)$$

$$= [-2 \ln L^*_{(C)}] - [-2 \ln L^*_{(A)}] \dots\dots\dots(14)$$

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

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

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 χ^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 fokal. Butir kedua berisi jawaban dari kelompok fokal yang tidak direspons oleh kelompok referensi. Keempat, menaksir kembali parameter, dan diperoleh harga χ^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

Hasil

Penelitian ini merupakan penelitian tahun kedua, yang merupakan penerapan dari hasil tahun pertama dalam rangka mengetahui sifat-sifat metode pendeteksian DIF di tahun pertama (2008) pada suatu perangkat tes. Langkah-langkah penerapan meliputi mengetahui dimensionalitas perangkat tes, mengetahui parameter butir perangkat tes berdasarkan teori tes klasik dan teori respons butir, kemudian mendeteksi DIF dengan pendekatan unidimensi dan multidimensi.

Penelitian tahun II difokuskan pada penelitian terapan hasil pengembangan tahun I. Ada beberapa data yang disiapkan untuk digunakan pada penelitian ini. Awalnya peneliti akan menggunakan data UN 2005, dan sudah diteliti UN mata pelajaran matematika 2005 ini mengukur dua dimensi. Pada proses analisisnya, menunjukkan bahwa data ini menunjukkan bahwa tes ini tidak ada butir yang memuat DIF. Selanjutnya dicoba data UN mata pelajaran matematika 2006, yang setelah dianalisis memuat 3 dimensi. Pada data ini, informasi yang ada adalah daerah asal. Dengan mengelompokkan siswa berdasarkan daerah asal, secara signifikan terbukti tidak ada DIF pada perangkat ini. Karena pemerintah (dinas pendidikan) tidak mengizinkan identitas peserta dikeluarkan, maka penelitian DIF dengan pengelompokkan berdasarkan jenis kelamin tidak dapat dilakukan. Selanjutnya digunakan data UN mata pelajaran matematika 2003.

Sebelum dilakukan analisis butir terkait dengan karakteristik UN 2003, terlebih dahulu dilakukan analisis faktor. Hasil analisis dengan uji Keiser-Meyer-Oldkin dan uji Bartlett diperoleh hasil sebagai berikut.

Tabel 3
Hasil Uji KMO dan Bartlett

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.963
Bartlett's Test of Sphericity	Approx. Chi-Square	20410.9
	df	17
		780

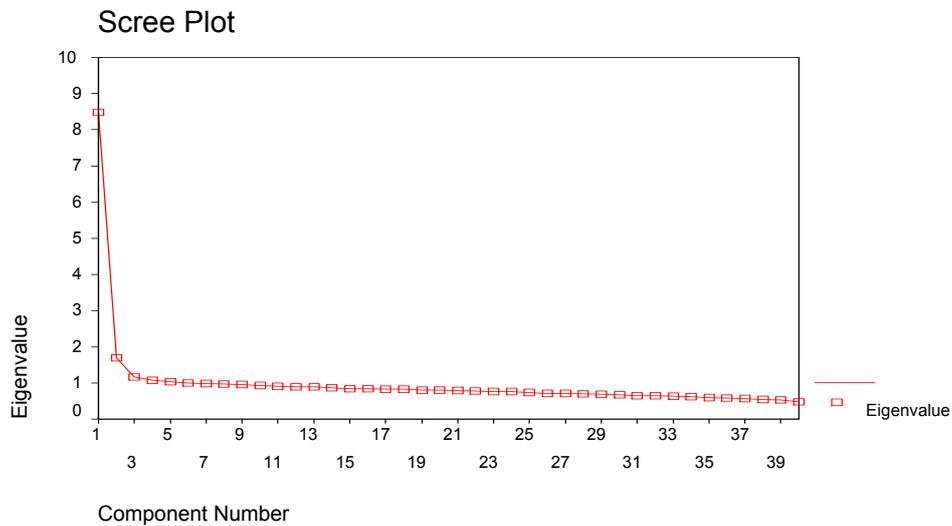
Sig.	.000
------	------

Hasil tersebut menunjukkan bahwa nilai KMO > 0,5, yang menunjukkan bahwa ukuran sampling untuk instrumen ini telah sesuai. Hasil Uji Bartlett juga menunjukkan bahwa nilai-p kurang dari 5%, yang menunjukkan bahwa butir-butir UAN matematika 2003 merupakan butir-butir yang saling independen. Selanjutnya, dengan memperhatikan nilai eigen yang lebih dari 1, dapat diketahui varians total yang dapat dijelaskan oleh perangkat UAN mata pelajaran matematika 2003. Hasil analisis menunjukkan bahwa ada 5 nilai eigen yang lebih dari 1, yang menunjukkan bahwa perangkat tes UAN matematika 2003 mengukur 5 faktor kemampuan matematika. Dengan menggunakan 40 butir UAN 2003 yang telah ada, varians total yang dapat dijelaskan hanya 33,654%. Hasil selengkapnya disajikan pada tabel berikut IV.5.

Tabel 4
Nilai Eigen dan Varians Total yang Dapat Dijelaskan

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.490	21.225	21.225	8.490	21.225	21.225	4.408	11.020	11.020
2	1.700	4.249	25.474	1.700	4.249	25.474	3.795	9.487	20.507
3	1.161	2.904	28.377	1.161	2.904	28.377	2.044	5.109	25.617
4	1.073	2.682	31.059	1.073	2.682	31.059	1.751	4.378	29.994
5	1.038	2.595	33.654	1.038	2.595	33.654	1.464	3.660	33.654

Nilai-nilai eigen hasil analisis ini dapat disajikan dengan Scree plot sebagai berikut.



Ga

Gambar 12. Scree Plot Nilai Eigen Hasil Analisis UAN Matematika 2003

Berdasarkan gambar 12 tersebut, nampak bahwa hanya ada 2 dimensi yang dominan dari 5 faktor yang ditemukan dengan analisis faktor.

B. Parameter Butir Perangkat Tes UN 2003

2. Hasil Analisis Butir yang Menggunakan Teori Respons Butir

2. Teori Respons Butir Bidimensi

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 8.

Tabel 8. Parameter Butir UN 2003 mapel Matematika 2 dimensi

Butir	C	b	a1	a2	Butir	c	b	a1	a2
1	0.26	0.644	0.834	0.109	21	0.202	0.918	1.795	0.525
2	0.155	-0.243	1.066	0.129	22	0.177	0.202	1.533	0.664
3	0.207	0.518	0.824	0.146	23	0.208	-0.162	0.983	0.383
4	0.1	-1.487	0.958	0.084	24	0.153	-0.319	0.944	0.041
5	0.27	0.348	1.204	-0.048	25	0.272	0.482	1.436	0.096
6	0.155	1.687	1.496	-0.31	26	0.351	0.413	1.27	0.163
7	0.257	1.416	0.838	0.129	27	0.325	-1.389	1.2	0.441
8	0.054	-0.322	1.448	-0.163	28	0.171	1.186	1.352	0.03

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
13	0.19	-1.366	1.814	0.101	33	0.458	-2.632	1.502	0.964
14	0.305	-0.29	1.472	-0.157	34	0.5	-1.307	1.19	0.526
15	0.082	0.346	0.871	-0.15	35	0.224	-0.666	0.785	0.378
16	0.188	1.792	0.787	-0.011	36	0.315	-0.594	0.811	0.363
17	0.319	-1.266	1.348	0.541	37	0.257	-1.184	0.905	0.54
18	0.202	-1.716	1.179	0.458	38	0.143	0.167	0.684	0.212
19	0.28	-1.943	1.547	-0.166	39	0.115	-0.526	0.875	0.194
20	0.231	-1.805	1.396	-0.109	40	0.226	0.179	1.093	0.472

Seperti halnya pada analisis butir unidimensi, pada analisis butir dengan pendekatan teori respons butir multidimensi dilakukan analisis terpisah kelompok wanita dan pria. Hasilnya disajikan pada tabel 9.

Tabel 9. 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 analisis DIF baik dengan pendekatan teori respons butir Unidimensi maupun berdasarkan Teori Respons Butir Multidimensi, sekaligus analisis signifikansinya dan kemudian dibandingkan.

D. Mengetahui Muatan DIF dengan IRT Bidimensi

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 yang dibangkitkan. Menguji signifikansi dilakukan dengan mengetahui perbedaan probabilitas dari sejumlah kemampuan yang dibangkitkan dan dengan menggunakan perbandingan likelihood. Hasil pada analisis dengan model bidimensi disajikan berikut ini.

1. Mengestimasi Indeks DIF

a. Estimasi Indeks dengan Perbedaan Volume di Bawah Kurva

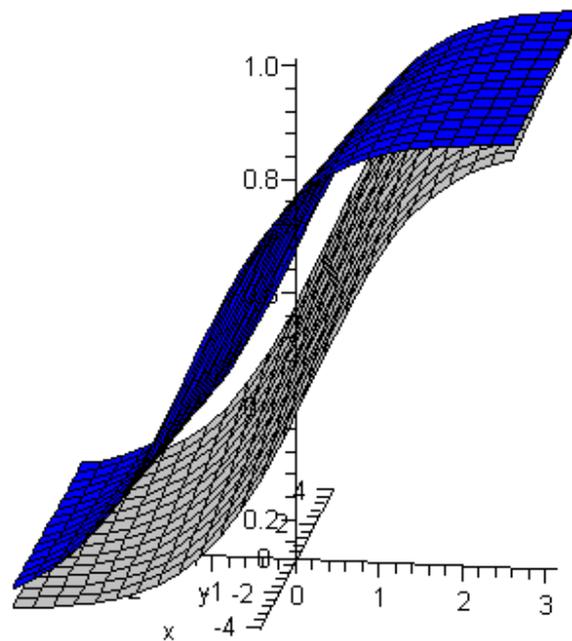
Untuk dapat menentukan hubungan antara permukaan karakteristik butir pada kelompok laki-laki dan kelompok perempuan, terlebih dahulu digambarkan kedua kurva tersebut pada ruang XYZ yang sama. Menggambarkan permukaan karakteristik

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 15.

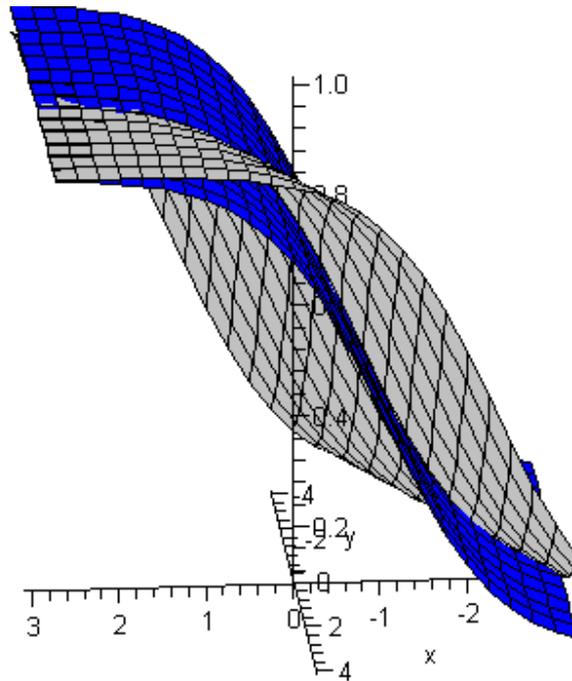
Tabel 15. Hasil Estimasi Indeks dengan Perbedaan Volume di Bawah Kurva

Butir	Hubungan Kedua Kurva	UPD	SPD
2	Tidak Berpotongan	2.375955194	0.319583065
3	Berpotongan	1.334156436	0.478482467
4	Berpotongan	2.380962549	0.340525261
5	Berpotongan	0.467709554	0.713328032
8	Berpotongan	3.050742576	0.699450513
9	Berpotongan	0.244860156	0.623609876
10	Berpotongan	1.560516793	0.104449195
11	Berpotongan	-0.05155445329	1.426478421
12	Berpotongan	4.15575866	0.566567029
18	Berpotongan	1.646013793	0.193323672
19	Berpotongan	-0.4881642811	2.633521896
20	Berpotongan	-0.4974348705	0.045147381
21	Berpotongan	0.591543786	2.222659109
22	Berpotongan	-1.1802804256	0.967730874
23	Berpotongan	0.658048647	0.038947217
24	Berpotongan	1.538900703	0.483387238
25	Berpotongan	0.623320644	0.054235151
26	Berpotongan	-0.07561402470	0.014026291
27	Berpotongan	1.333487292	0.686442311
29	Berpotongan	0.497429142	0.160806918
30	Tidak Berpotongan	6.285611838	1.326563291
31	Berpotongan	0.144309876	0.039878971
32	Berpotongan	-0.3964650968	1.069770056
33	Berpotongan	-0.08017496494	0.35498359
36	Berpotongan	0.261709532	0.202554996
37	Berpotongan	1.618482313	0.183737498
39	Berpotongan	1.563864728	1.226702913
40	Berpotongan	-0.1949786409	0.11756055

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 13 dan Gambar 14.



Gambar 13. Permukaan Karakteristik Butir Kelompok Laki-Laki (Permukaan Atas) Dan Kelompok Perempuan (Permukaan Bawah) Pada Butir Soal Nomor 30 (DIF Uniform)



Gambar 14. Permukaan Karakteristik Butir Kelompok Laki-Laki (Gelap) Dan Kelompok Perempuan (Abu-Abu) Pada Butir Soal Nomor 39 (DIF Nonuniform)

b. Esimasi dengan Perbedaan Peluang

Indeks DIF yang ditentukan dengan perbedaan peluang yang berupa UPD dan SPD ditentukan dengan bantuan EXCEL, setelah terlebih dahulu dibangkitkan 3000 kemampuan peserta dengan menggunakan model distribusi kemampuan dari peserta UN mata pelajaran matematika. Hasil estimasi disajikan pada Tabel 15.

Tabel 15. Hasil Estimasi Indeks DIF dengan Perbedaan Peluang

Butir	SPD	UPD	Butir	SPD	UPD
2	0.073215	0.006718	23	0.019558	0.000635
3	0.029485	0.001067	24	0.050342	0.003765
4	0.053291	0.004842	25	0.013035	0.00222
5	0.013816	0.00105	26	-0.00632	0.000457
8	0.106962	0.014457	27	0.070779	0.006547
9	0.008878	0.003199	29	0.007876	0.001288
10	0.046933	0.002706	30	0.064111	0.004906
11	0.019345	0.003265	31	0.026645	0.001309
12	0.064239	0.005772	32	0.011907	0.000892
18	0.041359	0.002262	33	-0.01986	0.00336
19	0.024011	0.003753	36	-0.00286	0.000369
20	-0.00313	0.000239	37	0.001111	0.00098
21	0.006216	0.00576	39	0.044907	0.004137
22	-0.00078	0.002535	40	-0.0029	0.000256

Mencermati hasil tersebut, dapat diperoleh bahwa hasil relatif kecil pada SPD maupun UPD. Hasil yang agak besar terjadi pada butir 2,4,10, 24, dan 30 yang dihitung dengan SPD.

2. Menguji Signifikansi

Menguji signifikansi indeks DIF dilakukan dengan perbandingan likelihood dan perbedaan probabilitas sampel.

a. Dengan Perbandingan Likelihood

Untuk dapat membandingkan likelihood antara model yang ditingkatkan ($G(A)$), dan pada model kompak ($G(C)$), terlebih dahulu diestimasi likelihoodnya. Estimasi dilakukan dengan bantuan TESTFACT dari Ssi. Hasil $G(C)$ - $G(A)$ selanjutnya dibandingkan dengan Khi-kuadrat tabel. Hasil selengkapnya disajikan pada Tabel 16.

Tabel 16. 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.

D. Pembahasan

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

luas/volume yang dihitung, semakin banyak titik kemampuan yang dibangkitkan, akan semakin mendekati luas/volume.

Indeks UPD dan SPD yang dihitung dengan perbedaan peluang, signifikansinya tergantung variabilitas dalam kelompok dan sangat tergantung pada varians inter kelompok. Hal ini menyebabkan hasil yang berbeda jika pengambilan anggota sub kelompok 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 antar variabel merupakan matriks singular, analisis yang dijalankan tidak konvergen. Akibat ketidakkonvergenan ini yakni tidak bisa teruji signifikansi DIFnya, karena nilai maksimum likelihoodnya tidak bisa diperoleh.

Pada model 2 dimensi hampir semua signifikan memuat DIF, 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- χ^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.

VI. PENUTUP

A. Kesimpulan

1. Metode Identifikasi DIF dengan pendekatan teori respons butir multidimensi yang dikembangkan tahun 2008 diterapkan pada data UN mata pelajaran matematika 2003.
2. Hasil estimasi indeks DIF (UPD dan SPD) hasilnya tidak terlalu berbeda, UPD ada yang negatif.
3. Perbandingan Likelihood lebih sensitif dalam menguji signifikansi muatan DIF dibandingkan dengan probabilitas sampling
4. Uji signifikansi muatan DIF dengan probabilitas sampling mudah dilakukan namun tidak terlalu sensitif, perbandingan likelihood lebih sensitif, namun lebih rumit pelaksanaannya.

B. Rekomendasi

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.ssssss

Kesimpulan

Ukuran DIF dapat diestimasi dengan menggunakan perbedaan probabilitas menjawab benar pada dua kelompok, yang disebut kelompok fokal dan kelompok referensi. Pada data multidimensi, perbedaan ini dapat diketahui dengan menggunakan indeks volume sederhana yang diketahui dengan menggunakan integral ganda. Signifikansi dari ukuran DIF antara dua kelompok dapat diuji dengan menggunakan perbandingan likelihood antara model kompak yang memuat butir yang diteliti dan model keseluruhan yang memuat keseluruhan butir.

Pustaka

- Ackerman, T.A., et. al. (2003). Using multidimensional 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 Marcov 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/tamu/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.
- Spray, J.A., et al. (1990). Comparison of two logistic multidimensional item response theory models. *ACT Research Report Series*. United States Government.