

Pendeteksian Keberfungsian Butir Diferensial (*Differential Item Functioning, DIF*) Menggunakan Indeks Perbedaan Probabilitas pada Data Politomus dengan Model *Generalized Partial Credit Model (GPCM)*

Oleh :
Kana Hidayati & Heri Retnawati
(Jurusan Pendidikan Matematika FMIPA UNY)

Abstrak

Salah satu syarat penyelenggaraan tes yakni tes bersifat adil. Tes bersifat adil jika tidak ada perbedaan peluang menjawab benar pada suatu kemampuan yang sama dari kelompok yang berbeda. Jika terdapat perbedaan peluang menjawab benar dari dua kelompok, dikatakan tes/butir memuat bias. Pada butir politomus dengan penskoran model *Generalized Partial Credit Model (GPCM)*, adanya muatan DIF dapat diketahui salah satunya dengan *probability difference indices*. Signifikansi ukuran ini dapat diketahui dengan *empirical sampling distribution for DIF indices*, yang dilakukan dengan membagi secara acak kelompok Focal (F) dan kelompok Reference (R) menjadi 2 kelompok, misalnya F1 dan F2, dan R1 dan R2, kemudian menghitung *probability difference indices* antara F1 dan F2, dan antara R1 dan R2. Pada tulisan ini dibahas tentang mengidentifikasi DIF pada IRT multidimensi dengan cara *probability difference indices* dan menguji signifikasinya dengan *empirical sampling distribution for DIF indices*.

Kata Kunci : DIF, Politomus, Indeks perbedaan probabilitas, GPCM

Latar belakang

Dalam suatu proses penilaian, idealnya tidak ada kesalahan baik kesalahan sistematis maupun kesalahan acak. Secara lebih khusus, seharusnya tidak ada kesalahan baik itu dari peserta tes, penyelenggaraan tes, maupun butir tes. Instrumen yang digunakan seharusnya valid, reliabel, stabil, dan adil. Adil diartikan bahwa tidak ada peserta yang dirugikan karena ketidakadilan tes tersebut. Jika ada perbedaan peluang menjawab benar menjawab benar suatu butir pada dua kelompok pada kemampuan yang sama, dikatakan butir ini memuat bias (*Differential Item Functioning, DIF*).

Namun kenyataannya tidak selalu demikian. Ada kalanya, skor hasil tes tidak memberikan informasi yang benar tentang kemampuan peserta tes. Mungkin dikarenakan informasi itu tidak menjangkau sampai ke besaran atau dimensi yang hendak diukur, atau mungkin hasil tes itu tercampur dengan besaran atau dimensi lain yang tidak dimaksudkan untuk diukur sehingga hasil tes itu rancu. Atau mungkin pula, pelaksanaan tes itu sendiri yang kurang layak sehingga menghasilkan informasi yang tidak benar.

Ada banyak metode untuk mengidentifikasi DIF yang dikembangkan oleh ahli psikometri menggunakan pendekatan teori respons butir, namun metode ini baru digunakan untuk identifikasi DIF pada penskoran dikotomi. Pada keadaan ini, analisis butir dengan pendekatan dikotomi tidak cocok lagi, dan akan menghasilkan kesalahan sistematis dan informasi yang menyesatkan.

Dengan memperhatikan karakteristik tes yang bersifat politomous, peneliti dapat menggunakan pendekatan teori respons butir dikotomi pada data politomus. Teori ini dapat digunakan untuk menganalisis butir suatu tes, termasuk mengidentifikasi DIF. Pada tulisan ini, akan dibahas pendeteksian DIF menggunakan indeks perbedaan probabilitas dua kelompok dan menguji signifikansinya dengan distribusi sampling pada data empiris.

Pembahasan

Dalam teori respons butir, hubungan antara peluang menjawab benar dengan skala kemampuan dan parameter butir dinyatakan dengan persamaan matematis yang disebut dengan model. Salah satu model matematis dengan penskoran politomus adalah *Generalized Partial Credit Model* (GPCM). GPCM menurut Muraki (1999) merupakan bentuk umum dari PCM, yang dinyatakan dalam bentuk matematis, yang disebut sebagai fungsi respons kategori butir sebagai berikut.

$$P_{jh}(\theta) = \frac{\exp \sum_{v=0}^h Z_{jr}(\theta)}{\sum_{e=0}^{m_j} \exp \left[\sum_{v=0}^e Z_{jr}(\theta) \right]}, \quad k=0,1,2,\dots,m_j$$

.....(1)

dan

$$Z_{jh}(\theta) = Da_j(\theta - b_{jh}) = Da_j(\theta - b_j + d_h), \quad b_{j0} = 0$$

.....(2)

Dengan

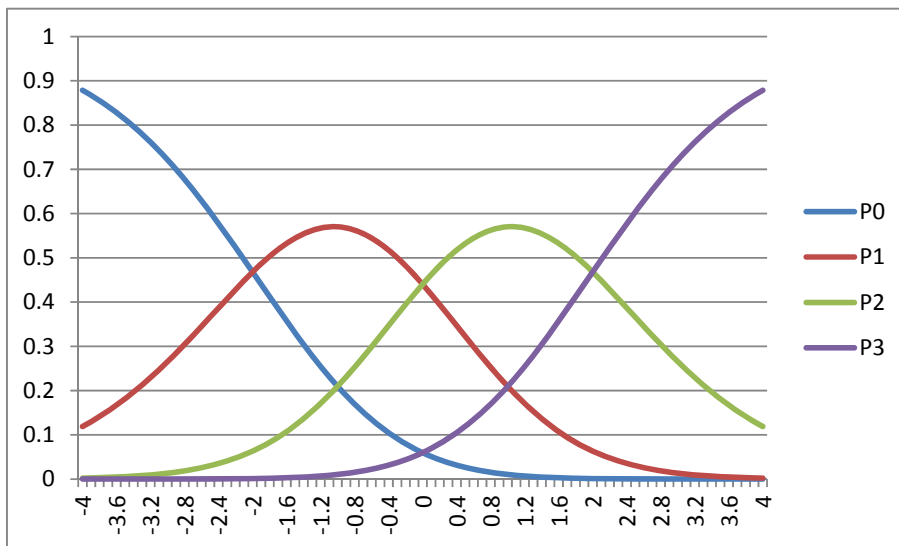
$P_{jk}(\theta)$: probabilitas peserta berkemampuan θ memperoleh skor kategori k pada butir j,

- θ : kemampuan peserta,
- a_j : indeks daya beda butir j,
- b_{jh} : indeks kesukaran kategori k butir j,
- b_j : indeks kesukaran lokasi butir j (parameter butir lokasi)
- d_k : parameter kategori k,
- m_j+1 : banyaknya kategori butir j, dan
- D : faktor skala (D=1.7)

Parameter b_{jh} oleh Master dinamai dengan parameter tahap butir. Parameter ini merupakan titik potong antara kurva $P_{jk}(\theta)$ dengan $P_{jk-1}(\theta)$. Kedua kurva hanya berpotongan di satu titik pada skala θ (van der Linden & Hambleton, 1997).

$$\begin{aligned} &\text{Jika } \theta = b_{jk}, \text{ maka } P_{jk}(\theta) = P_{jk-1}(\theta) \\ &\text{Jika } \theta > b_{jk}, \text{ maka } P_{jk}(\theta) > P_{jk-1}(\theta) \\ &\text{Jika } \theta < b_{jk}, \text{ maka } P_{jk}(\theta) < P_{jk-1}(\theta), \quad K=1,2,3,\dots,m_j \end{aligned} \tag{3}$$

GPCM diformulasikan berdasarkan asumsi bahwa setiap probabilitas memilih kategori ke-k melampaui kategori ke-(k-1) dibangun oleh model dikotomi. P_{jk} merupakan probabilitas khusus memilih kategori ke-k dari $m_j + 1$ kategori. Hubungan probabilitas menjawab benar untuk tiap kemampuan θ disajikan dalam grafik kurva karakteristik butir yang disebut dengan *Categorical Response Function* (CRF) (du Toit, 2003). Grafik CRF pada 4 kategori disajikan pada Gambar 1. Parameter butir untuk Gambar 1 tersebut yaitu daya pembeda (a) sebesar 1,0 dan tingkat kesulitan pada kategori menjawab -2,0, 0,0 dan 2,0



Gambar 1. Grafik CRF pada 4 kategori

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. 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*.

Bias butir atau DIF (*differential item functioning*) didefinisikan sebagai perbedaan peluang menjawab benar dari dua kelompok peserta yang dinamai dengan kelompok Fokal dan kelompok Referensi (Angoff, 1993; Lawrence, 1994; Hambleton & Rogers, 1995). Dalam teori respons butir unidimensi, DIF dinyatakan dengan peluang menjawab benar pada kelompok Referensi dikurangi peluang menjawab benar pada kelompok Fokal. Perbedaan peluang, DIF dinyatakan dengan indeks.

Indeks tersebut menyatakan sifat bias butir sebagai perbedaan peluang menjawab benar dari suatu tes yang diberikan kepada peserta yang kemampuannya sama, namun grupnya berbeda. Camilli dan Shepard (1994) secara matematis menyatakan perbedaan peluang sebagai

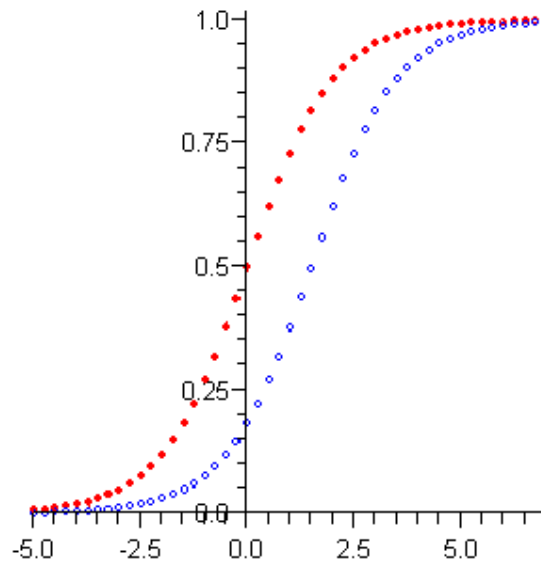
$$\Delta P_j = P_k(\theta_j) - P_p(\theta_j) \dots\dots\dots (4)$$

dengan ΔP_j merupakan perbedaan peluang pada peserta $P_k(\theta_j)$ j dari grup p ; $P_k(\theta_j)$ merupakan peluang menjawab benar peserta ke j dari grup k dengan kemampuan θ berdasarkan parameter butir pada skala kelompok k ; dan $P_p(\theta_j)$ merupakan peluang menjawab benar peserta ke j dari kelompok P dengan kemampuan θ berdasarkan skala parameter butir skala kelompok P .

Indeks perbedaan probabilitas disebut *Signed Probability Difference* (SPD) untuk memantau θ dinyatakan sebagai:

$$SPD - \theta = \frac{\sum_{j=1}^{n_p} \Delta P(\theta_j)}{n_p} \dots\dots\dots (5)$$

Persamaan digunakan untuk mengetahui perbedaan probabilitas dari dua kelompok yang kurva karakteristiknya tidak berpotongan seperti dinyatakan pada Gambar berikut.



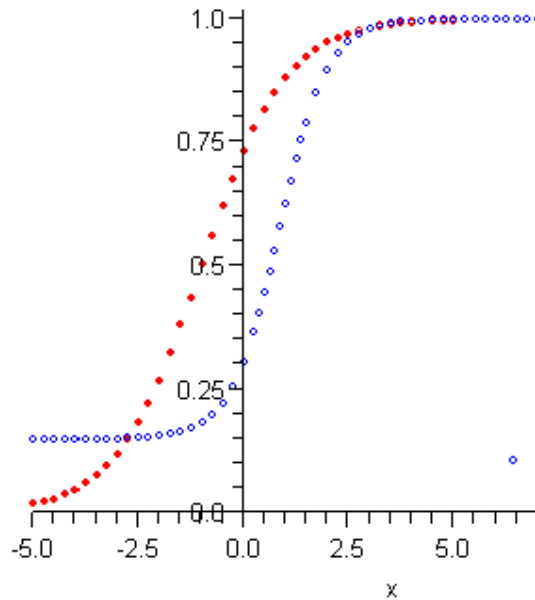
Gambar 2. Permukaan Karakteristik Butir untuk Menentukan Perbedaan Probabilitas dari Dua Kelompok (*DIF Uniform*)

Untuk butir yang kurva karakteristiknya saling berpotongan pada kedua grup, ukuran DIF dinyatakan dengan *Unsigned Probability Difference (UPD)* untuk memantau kemampuan θ dalam persamaan 6.

$$(6) \quad UPD - \theta = \sqrt{\frac{\sum_{j=1}^{n_p} (\Delta P(\theta_j))^2}{n_p}} \dots\dots\dots$$

Dengan n_p merupakan banyaknya peserta pada grup P , $\Delta P_p(\theta_j)$ perbedaan peluang menjawab benar peserta ke- j dari kelompok P , $SPD-\theta$ merupakan perbedaan probabilitas bertanda (*signed probability difference*) dicontrol oleh θ dari peserta ke- j , dan $UPD-\theta$ merupakan perbedaan probabilitas tidak bertanda (*unsigned probability difference*) dicontrol oleh θ untuk peserta ke- j (Camilli dan Shepard, 1994). Ukuran $SPD-\theta$ dapat positif dan negative dapat saling menghilangkan. $UPD-\theta$ merupakan ukuran kumulatif, sehingga tidak saling menghilangkan perbedaan kurva karakteristik. Ukuran DIF yang positif menyatakan bahwa kelompok P berada pada kondisi yang

tidak diuntungkan. Indikasi apakah ada perpotongan kurva karakteristik, jika indeks UPD- θ melebihi indeks SPD- θ . Perbedaan kecil antara indeks UPD- θ dan indeks SPD- θ secara praktis menunjukkan perpotongan antara dua kurva karakteristik butir tidak signifikan.



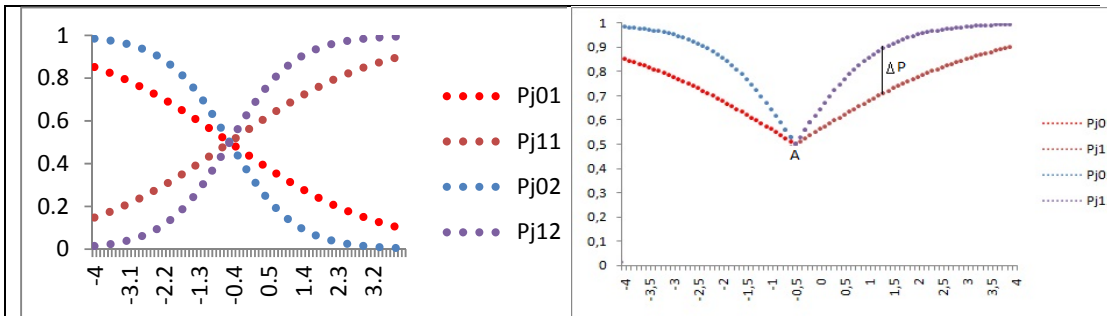
Gambar 3. Karakteristik Butir untuk Menentukan Perbedaan Probabilitas dari Dua Kelompok (*DIF Non-Uniform*)

Dengan menggunakan definisi DIF, perbedaan probabilitas antara kelompok Referensi dan kelompok Focal untuk menjawab dengan benar, dapat digunakan untuk mengidentifikasi DIF dalam teori respons butir politomus. Indeks perbedaan probabilitas untuk mengidentifikasi DIF dalam suatu butir dengan penskoran politomus dengan 2 kategori dapat dihitung dengan memperhatikan kurva karakteristik pada masing-masing kategori. Misalnya pada kurva karakteristik butir yang memuat DIF *uniform*, kurva pada kedua kategori baik untuk kelompok referensi maupun untuk kelompok focal berpotongan di titik A (gambar Untuk kurva karakteristik, dan gambar ...untuk menghitung selisih peluang). Jika selisih probabilitas dinyatakan dengan dinyatakan ΔP_j dan pada titik A adalah siswa ke n_A dengan maka

$$\Delta P_j = P_k(\theta_j) - P_p(\theta_j) \dots\dots\dots (7)$$

$$SPD - \theta = \frac{\sum_{j=1}^{n_p} \Delta P(\theta_j)}{n_p} = \frac{\sum_{j=1}^{n_A-1} \Delta P(\theta_j) + \sum_{j=n_A+1}^{n_p} \Delta P(\theta_j)}{n_p}$$

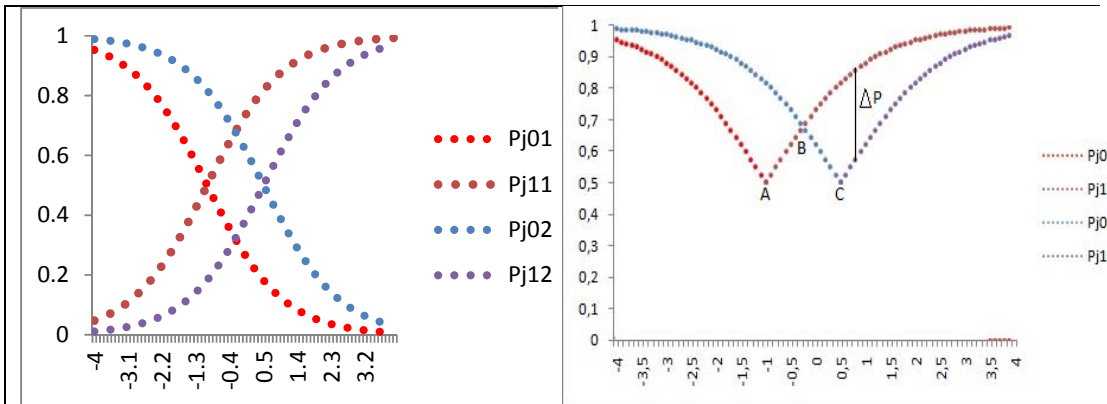
$$UPD - \theta = \sqrt{\frac{\sum_{j=1}^{n_p} (\Delta P(\theta_j))^2}{n_p}} = \sqrt{\frac{\sum_{j=1}^{n_A-1} (\Delta P(\theta_j))^2 + \sum_{j=n_A+1}^{n_p} (\Delta P(\theta_j))^2}{n_p}} \dots\dots\dots (8)$$



Gambar 4.a. Kurva Karakteristik Butir pada Kelompok R dan F dengan 2 Kategori (Memuat *DIF uniform*)

Gambar 4.b. Selisih Probabilitas Menjawab Benar Butir pada Kelompok R dan F pada Butir 2 Kategori (Memuat *DIF-uniform*)

Demikian pula halnya dengan DIF pada butir politomus dengan tiga kategori, perlu diperhatikan skala kemampuan ketika kurva karakteristik antar kategori atau antar kelompok saling berpotongan, misalnya A, B, dan C.



Gambar 5.a. Kurva Karakteristik Butir pada Kelompok R dan F dengan 2 Kategori (Memuat *DIF-nonuniform*)

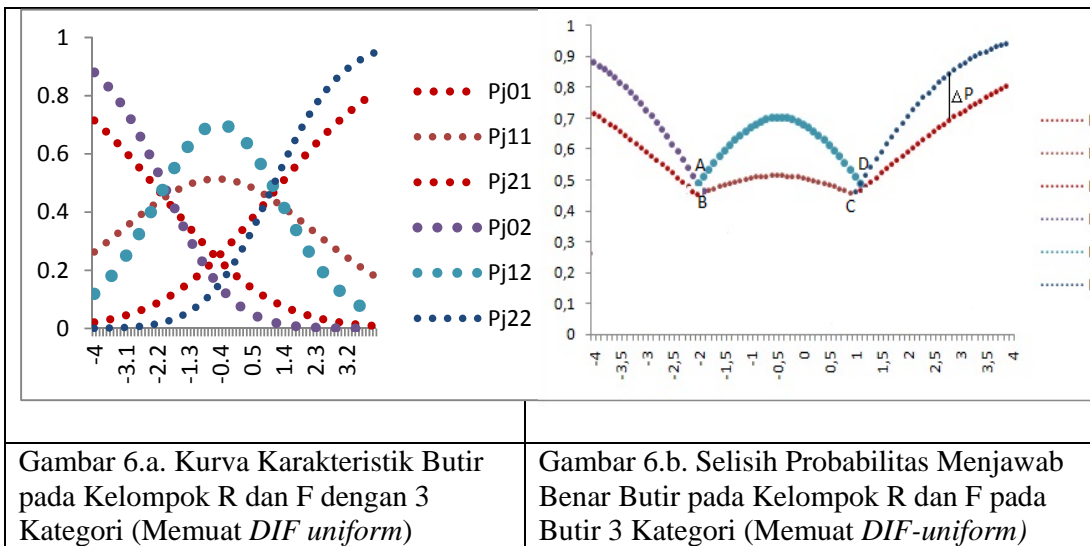
Gambar 5.b. Selisih Probabilitas Menjawab Benar Butir pada Kelompok R dan F pada Butir 2 Kategori (Memuat *DIF-nonuniform*)

$$UPD - \theta = \sqrt{\frac{\sum_{j=1}^{n_p} (\Delta P(\theta_j))^2}{n_p}}$$

$$= \sqrt{\frac{\sum_{j=1}^{n_A} ((P_{j02}(\theta_j) - P_{j01}(\theta_j))^2) + \sum_{j=n_A+1}^{n_B} ((P_{j02}(\theta_j) - P_{j11}(\theta_j))^2) + \sum_{j=n_B+1}^{n_C} ((P_{j02}(\theta_j) - P_{j11}(\theta_j))^2) + \sum_{j=n_C+1}^{n_p} ((P_{j02}(\theta_j) - P_{j11}(\theta_j))^2)}{n_p}}$$

.....(9)

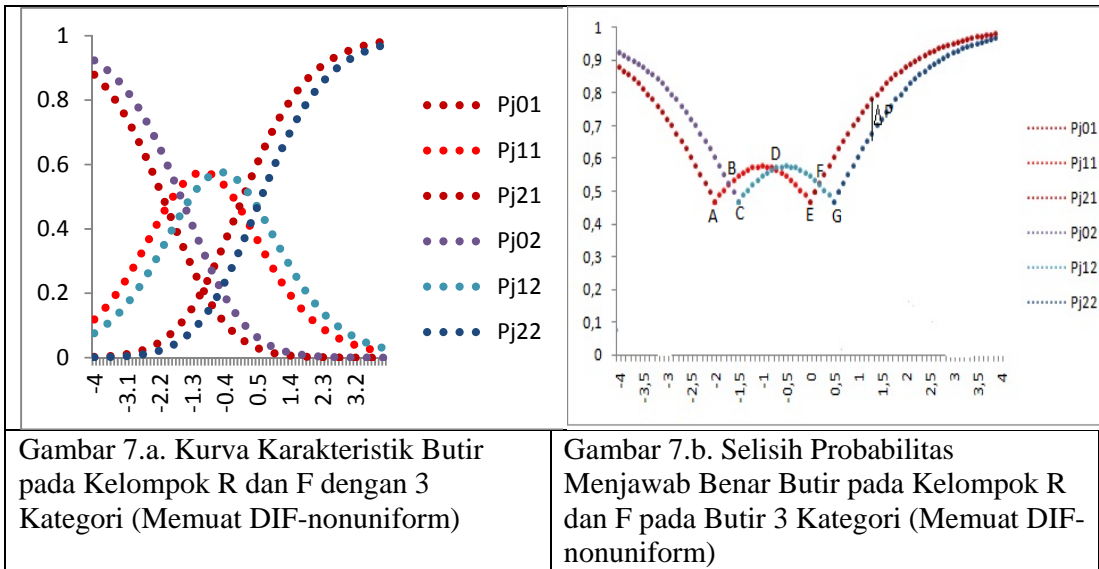
Demikian pula pada butir yang memuat DIF-uniform, daerah asal perpotongan antar kurva untuk tiap kategori pada tiap kelompok.



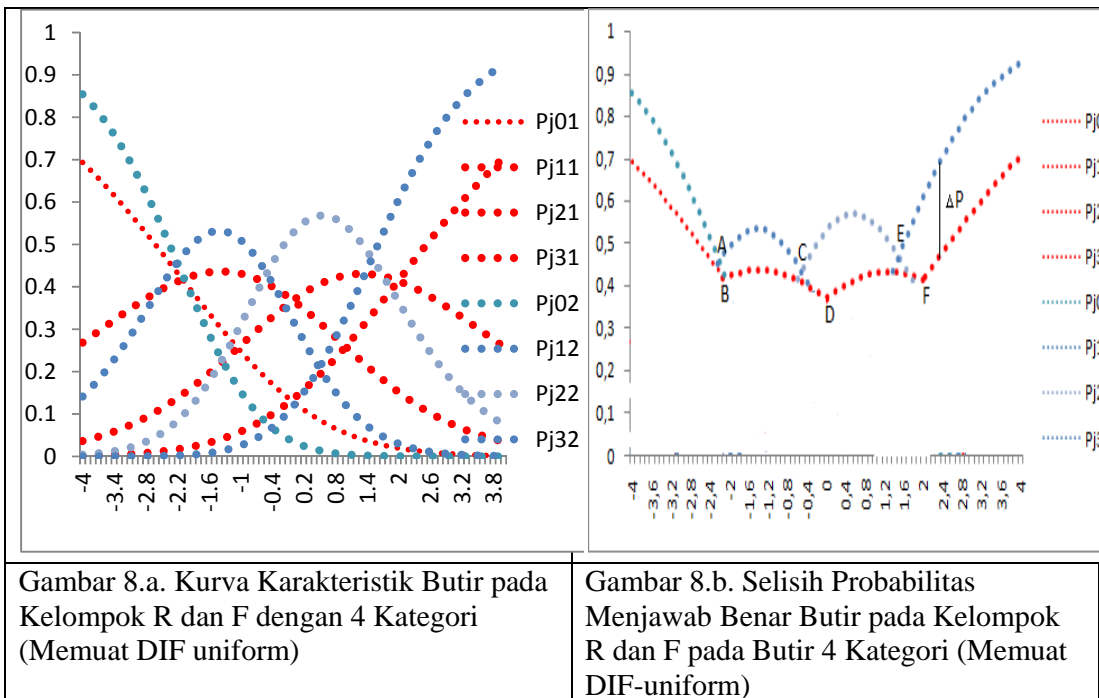
Untuk butir yang memuat DIF-uniform, daerah asal perpotongan antar kurva untuk tiap kategori pada tiap kelompok dan antar kelompok perlu memperoleh perhatian, yang kemudian menjadi dasar untuk menyusun formula mengestimasi indeks DIF.

$$SPD - \theta = \frac{\sum_{j=1}^{n_A} (P_{j01}(\theta_j) + \sum_{j=n_A+1}^{n_B} P_{j11}(\theta_j) + \sum_{j=n_B+1}^{n_p} P_{j21}(\theta_j) - \sum_{j=1}^{n_B} P_{j02}(\theta_j) - \sum_{j=n_B+1}^{n_C} P_{j12}(\theta_j) - \sum_{j=n_C+1}^{n_p} P_{j22}(\theta_j))}{n_p}$$

Untuk butir yang memuat DIF-nonuniform, indeks DIF diketahui dengan UPD yang diestimasi dengan

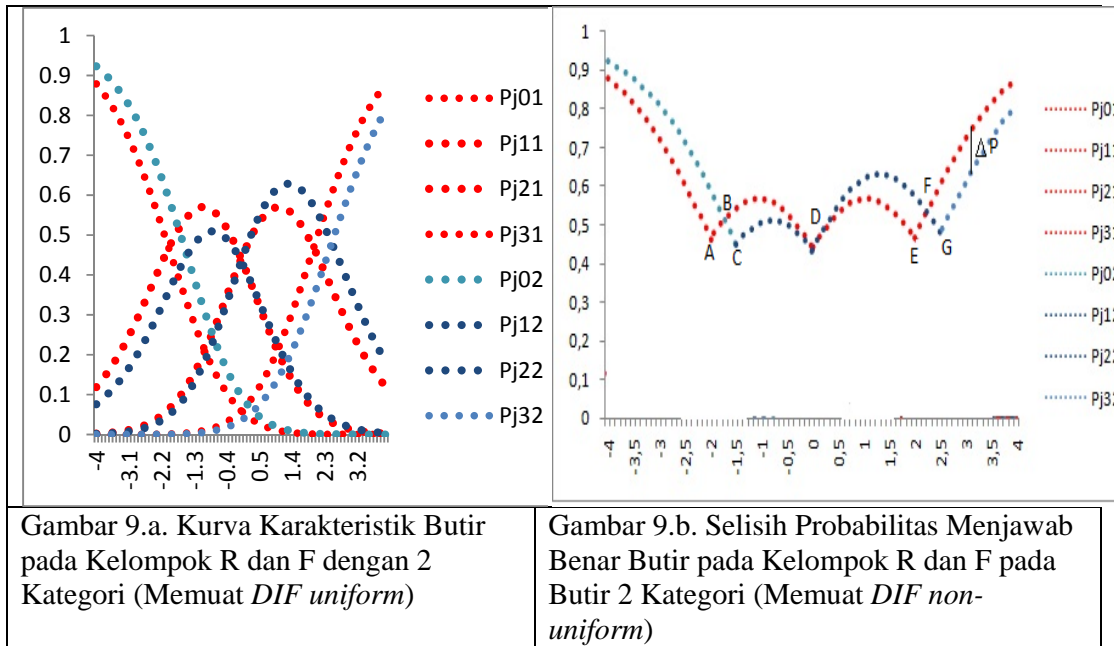


$$\text{UPD-}\theta = \frac{\sum_{j=1}^{n_a} (P_{j1}(\theta) - P_{j02}(\theta))^2 + \sum_{j=n_a+1}^{n_b} (P_{j01}(\theta) - P_{j12}(\theta))^2 + \sum_{j=n_b+1}^{n_c} (P_{j12}(\theta) - P_{j01}(\theta))^2 + \sum_{j=n_c+1}^{n_d} (P_{j12}(\theta) - P_{j11}(\theta))^2 + \sum_{j=n_d+1}^{n_e} (P_{j11}(\theta) - P_{j22}(\theta))^2 + \sum_{j=n_e+1}^{n_p} (P_{j21}(\theta) - P_{j22}(\theta))^2}{n_p} \quad (10)$$



$$SPD-\theta = \frac{\sum_{j=1}^{n_A} P_{j01}(\theta_j) + \sum_{j=n_A+1}^{n_C} P_{j11}(\theta_j) + \sum_{j=n_C+1}^{n_E} P_{j21}(\theta_j) + \sum_{j=n_E+1}^{n_p} P_{j31}(\theta_j) - \sum_{j=1}^{n_B} P_{j02}(\theta_j) - \sum_{j=n_B+1}^{n_D} P_{j12}(\theta_j) - \sum_{j=n_D+1}^{n_F} P_{j22}(\theta_j) - \sum_{j=n_F+1}^{n_p} P_{j32}(\theta_j)}{n_p} \dots (1)$$

1)



UPD-θ

$$UPD-\theta = \frac{\sum_{j=1}^{n_A} (P_{j01}(\theta_j) - P_{j02}(\theta_j))^2 + \sum_{j=n_A+1}^{n_C} (P_{j11}(\theta_j) - P_{j12}(\theta_j))^2 + \sum_{j=n_C+1}^{n_E} (P_{j21}(\theta_j) - P_{j22}(\theta_j))^2 + \sum_{j=n_E+1}^{n_p} (P_{j31}(\theta_j) - P_{j32}(\theta_j))^2}{n_p} \dots (12)$$

Signifikansi dari ukuran DIF dalam suatu butir dapat diketahui dengan distribusi sampling Empiris untuk indeks DIF (Camilli and Shepard, 1994). Dalam metode ini, kelompok Referensi dan Kelompok Focal keduanya secara acak random dipisah menjadi dua bagian, menjadi subsample R1 dan R2, F1 dan F2 dan mendapat perlakuan yang sama seperti kelompok Referensi dan kelompok Focal. Karena kedua subsample iini disusun secara acak, idealnya tidak ada muatan DIF. Indeks DIF yang ditunjukkan terjadi karena kedua grup menunjukkan kesalahan penyampelan. Nilai ekstrem dari analisis R1-R2 dan F1-F2 dipilih sebagai nilai kritis menandai sigifikansinya muatan DIF.

Kesimpulan

Dalam teori respons butir, keberadaan DIF pada data politomus model GPCM dapat diselidiki dengan indeks *Signed Probability Difference* yang dikontrol kemampuan (SPD- θ) jika kurva karakteristik dua kelompok berpotongan dan indeks *Unsigned Probability Difference* yang dikontrol kemampuan (UPD- θ) jika kurva karakteristik dua kelompok berpotongan. Signifikansi dari ukuran ini dapat diketahui dengan menggunakan distribusi sampling empiris untuk indeks DIF, dengan membagi acak kelompok Fokal (F) dan kelompok Referensi (R) menjadi dua sub kelompok, sebagai contoh F1 dan F2, dan R1 dan R2, kemudian menghitung perbedaan probabilitas antara F1 dan F2, dan antara R1 dan R2. Nilai ekstrem hasil analisis indeks DIF dari R1-R2 dan F1-F2 dipilih sebagai nilai kritis signifikansi DIF.

Daftar Pustaka

- 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.
- Camilli, G. & Shepard, L.A. (1994). *Methods for identifying biased test items, Vol.4*. London: Sage Publications, Inc.
- Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: SSI.
- Gronlund, N.E. & Linn, R.L. (1990). *Measurement and evaluation in teaching (6th ed)*. New York : Collier Macmillan Publishers
- 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.
- Muraki, E. (1999). New approaches to measurement. Dalam Masters, G.N. dan Keeves, J.P.(Eds). *Advances in measurement in educational research and assesment*. Amsterdam : Pergamon.
- Thisen, D. et al. (1993). Detection of differential item functioning using the parameters of item respons model. Dalam P.W. Holland & H. Wainer (Ed). *Differential Item Functioning*. Englewood Clifp, NJ : Lawrence Erlbaum. p.67-86.
- Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.