

METODE *LINKING* DAN BUTIR *FALSE POSITIVE* PADA PENDETEKSIAN *DIF* BERDASARKAN TEORI RESPONSI BUTIR

Wardani Rahayu
PPs UNJ
wardani9164@yahoo.com

Abstract

Tujuan dari penelitian ini adalah untuk mengetahui pengaruh metode *linking* pada sejumlah butir *false positif* dalam pendeteksian *DIF* berdasarkan teori response butir. Metode *linking* yang digunakan adalah metode rerata dan sigma (RS), metode tegar rerata dan sigma (TRS), metode kurva karakteristik (KK). Pendeteksian *DIF* dilakukan dengan menggunakan metode Lord's Chi-Square Test. Hasil pendeteksian *DIF* melalui metode *Lord's Chi-Square* ini didasarkan pada banyaknya butir *false positive* (FP) yang mempengaruhi keakuratan metode *linking*. Jumlah butir *false positif* dihitung dari butir yang sebelumnya diperkirakan tidak *DIF* berdasarkan *Likelihood Ratio Test*, akan tetapi setelah pendeteksian *DIF* menggunakan Lord's Chi-Square Test, ternyata ditemukan bahwa item mengandung *DIF*. Hasil penelitian menunjukkan bahwa metode kurva karakteristik, metode rerata dan metode sigma adalah lebih akurat daripada metode tegar rerata dan metode sigma. Namun, metode kurva karakteristik memiliki akurasi yang sama dengan metode rerata dan metode sigma.

Kata kunci: *metode linking, butir false positive, DIF*

LINKING METHOD AND FALSE POSITIVE ITEM ON DIF DETECTION BASED ON ITEM RESPONSE THEORY

Wardani Rahayu

Graduate School of Jakarta State University
wardani9164@yahoo.com

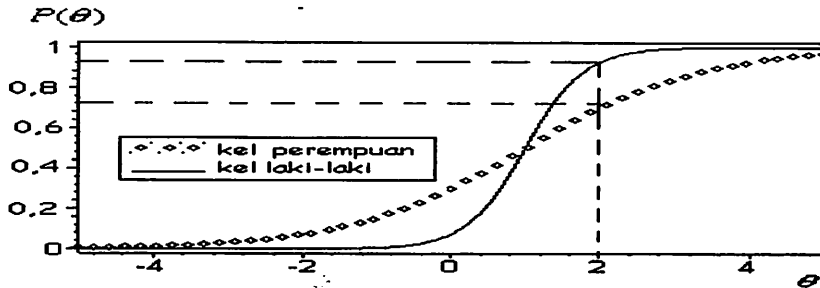
Abstract

The objective of this research was to investigate the effect of the linking methods on the number of false positive items in DIF detection based on item response theory. The linking methods which was compared was mean and sigma method, robust mean and sigma method, characteristic curve method. The method used to detect DIF was the Lord's Chi-Square Test. The achievement of this research was carried out through the calculation of the number of false positive items from the result of DIF detection using the the Lord's Chi-Square Test. The number of false positive items was calculated from the items which was formerly predicted not DIF based on Likelihood Ratio Test, but after the DIF detection using the Lord's Chi-Square Test, it was found out that the items were DIF. The results were that the characteristic curve method and mean and sigma method were more accurate than robust mean and sigma method. But the characteristic curve method had the same accuracy as mean and sigma method.

Kata kunci: *linking methods, false positive, dif*

Pendahuluan

Teori responsi butir membangun model yang menghubungkan parameter butir dengan parameter kemampuan peserta tes. Model yang dikembangkan berlaku untuk semua kelompok butir dan semua kelompok peserta tes yang memenuhi beberapa syarat tertentu. Model hubungan



Gambar 1. Dua Kurva Karakteristik Butir ke- j Model L2P

Gambar 1, absis menyatakan kemampuan peserta tes sedangkan ordinat menyatakan $P(\theta)$ atau probabilitas menjawab betul kemampuan peserta tes θ . Pada $\theta = 2$ didapat probabilitas menjawab betul kelompok laki-laki dan perempuan yang berbeda. Artinya, pada kemampuan peserta kelompok laki-laki dan perempuan sama dengan 2, mereka memiliki probabilitas menjawab betul yang tidak sama pada butir ke- j , sehingga butir ke- j dikatakan mengandung *DIF*.

Berdasarkan penjelasan di atas dapat disimpulkan bahwa dua kelompok peserta tes yang memiliki kemampuan sama memiliki probabilitas menjawab betul yang sama pada suatu butir maka dikatakan butir tersebut tidak mengandung *DIF*. Suatu butir tidak mengandung *DIF* secara langsung berkaitan dengan asumsi invarian dalam teori responsi butir.

Deteksi *DIF* pada suatu butir dilakukan dengan menggunakan analisis statistika. Ada beberapa metode yang dapat digunakan untuk mendeteksi *DIF* pada suatu butir di antaranya *Lord's Chi-Square*, *Likelihood Ratio Test*. Untuk mendeteksi suatu butir mengandung *DIF* atau tidak maka terlebih dahulu parameter butir dari dua kelompok peserta tes yang berbeda dilakukan penyamaan skala. Ini bertujuan agar parameter butir dari dua kelompok tersebut berada pada metrik yang sama. Metode penyamaan skala disebut dengan metode *linking*. Penyamaan skala dapat dilakukan secara terpisah antara dua kelompok peserta tes atau dapat

dilakukan secara simultan tergantung dari metode *DIF* yang akan digunakan.

Menurut Kim dan Cohen (1998:117) *linking* merupakan proses penyamaan skala dalam teori responsi butir yaitu mentransformasikan parameter butir dari metrik yang satu ke metrik yang lain. Data yang berupa respon peserta tes kedua kelompok diestimasi secara terpisah sehingga didapat estimasi parameter butir dari dua kelompok peserta tes tersebut. Selanjutnya, parameter butir tersebut ditransformasikan dari satu kelompok ke kelompok yang lain. Metode *linking* yang digunakan dalam penelitian ini adalah metode rerata dan sigma (RS), metode tegar rerata dan sigma (TRS), metode kurva karakteristik (KK).

Keakuratan metode *linking* dalam pendekteksian *DIF* berhubungan dengan banyaknya butir *false positive* (FP) yang muncul pada setiap metode *linking*. Menurut Kim dan Cohen (1992) *false positive* (FP) dalam pendekteksian *DIF* suatu butir pada perangkat tes terjadi apabila suatu butir tes yang diperkirakan tidak *DIF*, ternyata *DIF*. Keakuratan metode *linking* pada pendekteksian *DIF* melalui metode *Lord's Chi-Square* dilihat dari banyaknya butir *false positive* (FP) yang teridentifikasi pada setiap metode *linking*. Semakin sedikit butir *false positive* (FP) yang muncul pada metode *linking* maka disimpulkan bahwa salah satu metode *linking* tersebut paling akurat.

Penyamaan skala menggunakan metode RS hanya melibatkan taraf sukar butir. Rumus penyamaan skala ini mengandung koefisien *A* dan *K*. Koefisien *A* diperoleh dari perbandingan antara simpangan baku taraf sukar butir kelompok referensi dan kelompok fokal, sedangkan koefisien *K* diperoleh dari perbedaan rerata taraf sukar butir kelompok referensi, dan perkalian *A* dengan rerata taraf sukar butir kelompok fokal. Estimasi parameter butir kelompok fokal dan referensi dilakukan secara terpisah. Hasil transformasi parameter butir kelompok fokal dipengaruhi oleh nilai *A* dan *K*, sedangkan nilai *A* dan *K* dipengaruhi oleh nilai taraf sukar butir kelompok fokal dan kelompok referensi.

Penyamaan skala menggunakan metode TRS juga hanya melibatkan taraf sukar butir. Rumus penyamaan skala ini mengandung koefisien *A*

dan K . Koefisien A diperoleh dari perbandingan antara simpangan baku taraf sukar butir berbobot kelompok referensi dan kelompok fokal, sedangkan koefisien K diperoleh dari perbedaan rerata taraf sukar butir berbobot kelompok referensi dan perkalian A dengan rerata taraf sukar butir berbobot kelompok fokal. Taraf sukar butir berbobot kelompok fokal dan kelompok referensi pada butir ke- j diperoleh dari perkalian taraf sukar dengan bobot skala. Bobot skala butir ke- j memiliki nilai diantara 0 dan 1 sehingga taraf sukar butir berbobot kurang dari nilai taraf sukar butir yang belum diboboti. Estimasi parameter butir kelompok fokal dan kelompok referensi dilakukan secara terpisah.

Sementara metode KK yang perhitungan koefisien transformasi linear A dan K dilakukan serentak. Rumus penyamaan skala ini mengandung koefisien A dan K . Koefisien A dan K diperoleh

dengan cara meminimalkan fungsi $F = \frac{1}{M} \sum_{g=1}^M \xi_{gF}(\theta)$. Fungsi F

dipengaruhi oleh hasil estimasi taraf sukar butir, daya beda butir, dan taraf kebetulan menjawab betul pada butir serta kemampuan peserta tes. Hasil perhitungan koefisien A dan K dipengaruhi pula oleh hasil estimasi taraf sukar butir, daya beda butir, dan taraf kebetulan menjawab betul pada butir serta kemampuan peserta tes

Tujuan penelitian pada pendeteksian DIF melalui metode *Lord's Chi-Square* ini, untuk mengetahui banyaknya butir *false positive* (FP) yang mempengaruhi keakuratan metode *linking*. Di samping itu mengembangkan software STRING, BLINK dan LORDDIF yang digunakan untuk penyamaan skala dan pendeteksian DIF.

Metodologi Penelitian

Penelitian ini termasuk dalam penelitian eksperimen yang terdiri atas dua variabel yaitu variabel terikat dan variabel bebas. Variabel terikat dari penelitian ini adalah banyaknya butir FP dan variabel bebas adalah metode *linking* yaitu metode RS, TRS dan KK.

Data yang digunakan dalam penelitian ini berupa skor pekerjaan siswa. Skor hasil pekerjaan siswa berasal dari respons siswa terhadap perangkat tes Ujian Nasional SMP mata pelajaran matematika paket 3 untuk wilayah Kodya Jakarta Timur dan Jakarta Barat yang dilaksanakan pada tahun 2004.

Langkah-langkah penelitian ini dilakukan sebagai berikut pengambilan data berupa skor pekerjaan siswa SMP untuk wilayah Kodya Jakarta Timur dan Jakarta Barat di Pusat Penilaian dan Pengujian Pendidikan Departemen Pendidikan Nasional. Data utama yang digunakan berupa skor yang berbentuk nol (0) dan satu (1) dengan panjang perangkat tes 40 butir. Perangkat tes yang terdiri atas 40 butir dideteksi *DIF*nya dengan menggunakan metode *Likelihood Ratio Test*. Ini bertujuan untuk memperoleh deskripsi data tentang butir-butir mana yang mengandung *DIF* atau tidak. Jumlah respon peserta tes kelompok referensi dan kelompok fokal yang digunakan untuk mendeteksi *DIF* masing-masing 29242.

Pendeteksian *DIF* dengan menggunakan metode *Likelihood Ratio Test* dilakukan melalui dua tahap. Tahap pertama dengan menggunakan metode *all other* yaitu semua butir dijadikan jangkar untuk deteksi *DIF*. Tahap dua menggunakan empat butir bebas *DIF* untuk dijadikan jangkar. Menurut Wang dan Yeh (2003), tipe galat 1 yang dihasilkan metode *all-other* akan bersifat stabil apabila perangkat tes tidak terdapat butir-butir yang mengandung *DIF* atau maksimal hanya terdapat satu butir yang mengandung *DIF*. Tipe galat I akan stabil jika butir yang digunakan sebagai jangkar adalah butir yang bebas *DIF* dengan jumlah 4 butir. Bertambahnya jumlah jangkar bertambah pula tingkat kestabilannya. Deteksi *DIF* dengan metode ini menggunakan program BILOG.

Jumlah respon peserta tes kelompok referensi dan kelompok fokal masing-masing 29242. Diambil 1000 respon peserta tes sebanyak 384 kali secara acak untuk masing-masing kelompok referensi dan kelompok fokal. Pengambilan dilakukan menggunakan program MINITAB. Selanjutnya, mengestimasi parameter butir dan parameter kemampuan peserta tes gabungan peserta kelompok referensi dan fokal pada setiap replikasi. Jumlah respon peserta yang digunakan untuk mengestimasi parameter butir

dan parameter kemampuan peserta tes sebanyak 2000 terdiri dari 1000 respon kelompok referensi dan 1000 respon kelompok fokal. Estimasi dilakukan dengan menggunakan program BILOG. Dari hasil estimasi ini dicermati butir-butir yang cocok model, ternyata terdapat 56 replikasi yang mengandung 25 butir dengan nomor sama yang cocok model dari 384 replikasi.

Respon peserta tes yang tersaji untuk 25 butir yang cocok model kemudian diestimasi parameter butir dan parameter peserta tes, variansi kovariansi butir, fungsi informasi tes kelompok referensi dan kelompok fokal secara terpisah. Estimasi ini dilakukan dengan menggunakan program BILOG. Selanjutnya, menyamakan skala daya beda butir a , taraf sukar butir b , dan tingkat kebetulan menjawab betul pada butir c dari kelompok fokal ke kelompok referensi dengan menggunakan tiga metode *linking* yang berbeda yaitu RS, TRS dan KK. Penyamaan skala dengan metode RS dan TRS menggunakan program BLINK. Penyamaan skala dengan metode KK menggunakan program EQUATE 2.

Dilanjutkan deteksi *DIF* dengan metode *Lord's Chi-Square* dan mengidentifikasi butir-butir FP, kemudian menghitung banyaknya butir FP yang bersesuaian dengan metode *linking* yang digunakan yaitu RS, TRS, dan KK. Deteksi *DIF* menggunakan program LORDDIF.

Penelitian ini mengembangkan software program STRING, BLINK dan LORDDIF. STRING adalah suatu program komputer yang digunakan untuk merapikan data menurut kolom secara simultan yang tersaji di Notepad. BLINK merupakan software yang digunakan untuk penyamaan skala dengan metode RS dan TRS. LORDDIF adalah program komputer yang digunakan untuk mendeteksi suatu butir *DIF* atau tidak berdasarkan metode *Lord's Chi-Square*. Model yang digunakan dalam program ini adalah logistik.

Teknik Analisis Data

1. Uji Persyaratan Analisis.

Uji kehomogenan varians pada butir *FP* hasil deteksi *DIF* yang penyamaan skalanya menggunakan metode *TRS*, *RS* dan *KK* dengan menggunakan uji *Bartlett*. Hasil pengujian *Bartlett* didapat kehomogenan varians tidak terpenuhi. Dengan demikian, pengujian hipotesis menggunakan statistik nonparametrik

2. Uji Perbedaan Rerata dan Uji perbandingan berganda

Pengujian hipotesis dalam penelitian ini dilakukan melalui dua tahap yaitu pertama uji perbedaan rerata dan kedua uji perbandingan berganda. Uji perbedaan rerata banyaknya butir *FP* dari tiga metode *linking* dalam pendeteksian *DIF* menggunakan uji *Kruskal Wallis*. Uji perbedaan rerata dengan uji *Kruskal Wallis* digunakan untuk menguji adanya perbedaan pengaruh perlakuan yang dicobakan. Karena ada dua rerata yang berbeda maka dilakukan uji lanjut yaitu uji perbandingan berganda.

Hasil Penelitian

Pendeteksian *DIF* dengan menggunakan metode *Likelihood Ratio Test* dilakukan melalui dua tahap. Tahap pertama dengan menggunakan metode *all other* yaitu semua butir dijadikan jangkar untuk deteksi *DIF*. Tahap dua menggunakan empat butir bebas *DIF* untuk dijadikan jangkar. Hasil pendeteksian *DIF* pada tahap satu adalah 32 butir *DIF* dan 8 butir tidak *DIF*. Butir-butir yang tidak mengandung *DIF* adalah butir ke- 9, 11, 12, 17, 18, 29, 37, dan 23, sedangkan butir yang lain merupakan butir *DIF*.

Butir ke-23, 11, 18 dan 17 digunakan sebagai jangkar dalam pendeteksian *DIF* pada tahap dua, sehingga jumlah butir yang diselidiki *DIF* atau tidaknya pada tahap dua sebanyak 36 butir. Hasil deteksi *DIF* pada tahap dua adalah 21 butir *DIF* dan 15 butir tidak *DIF*. Hasil deteksi *DIF* 40 butir pada perangkat tes Ujian Akhir Nasional SMP mata pelajaran matematika paket 3 dengan menggunakan metode *Likelihood Ratio Test* dapat dilihat pada Tabel 1.

Tabel 1. Butir-butir *DIF* dan Tidak *DIF* Hasil Deteksi *DIF* Tahap 2 dengan Metode *Likelihood Ratio Test*

Butir-Butir Tidak Mengandung <i>DIF</i>	Butir-Butir Mengandung <i>Dif</i>
9, 11, 12, 13, 17, 18, 23, 27, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	1, 2, 3, 4, 5, 6, 7, 8, 10, 14, 15, 16, 19, 20, 21, 22, 24, 25, 26, 28, 30
19 butir	21 butir

Butir-butir pada perangkat tes Ujian Akhir Nasional SMP mata pelajaran matematika paket 3 yang dideteksi dengan metode *Lord's Chi-Square* sebanyak 25 butir. Sebelumnya butir-butir ini telah dideteksi dengan menggunakan metode *Likelihood Ratio Test*. Hasil deteksi *DIF* 25 butir dengan menggunakan metode *Likelihood Ratio Test* adalah 11 butir mengandung *DIF* dan 14 butir tidak mengandung *DIF*.

Penelitian ini berhubungan dengan butir *FP* yaitu apabila butir yang dinyatakan tidak mengandung *DIF* ketika dideteksi dengan metode *Likelihood Ratio Test*, ternyata mengandung *DIF* setelah dideteksi dengan metode *Lord's Chi-Square*. Oleh karena itu, fokus penelitian ini hanya tertuju pada 14 butir tidak mengandung *DIF* sebelum dideteksi dengan metode *Lord's Chi-Square*. Keempat belas butir tidak mengandung *DIF* itu adalah butir ke- 9, 11, 12, 13, 18, 23, 27, 29, 31, 33, 34, 37, 38 dan 40.

Hasil pengujian dengan *Kruskal Wallis* menunjukkan ada perbedaan antara rerata banyaknya butir *FP* pada metode RS, rerata banyaknya butir *FP* pada metode TRS, dan rerata banyaknya butir *FP* pada metode KK pada pendeteksian *DIF* melalui metode *Lord's Chi-Square*. Dengan demikian, ketiga metode *linking* yang dicobakan tidak memberikan pengaruh yang sama, maka ada pertanyaan berikut yang perlu dijawab adalah metode *linking* manakah yang memberi pengaruh yang berbeda. Oleh karena itu, dilakukan pengujian lanjutan yaitu uji perbandingan berganda.

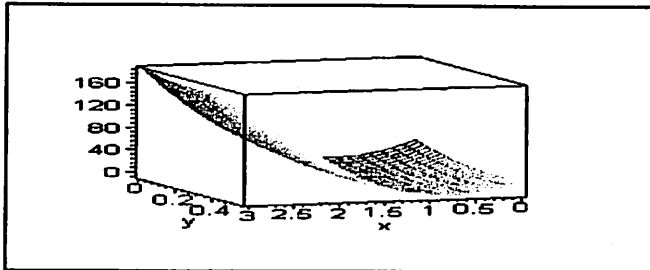
Hasil pengujian perbandingan berganda menunjukkan bahwa banyaknya butir *FP* pada metode RS kurang dari banyaknya butir *FP* pada metode TRS, banyaknya butir *FP* pada metode RS sama dengan banyaknya butir *FP* pada metode kurva karakteristik (KK), dan banyaknya butir *FP* pada metode TRS lebih dari banyaknya butir *FP* pada metode

KK. Keakuratan metode *linking* dilihat dari semakin sedikit butir FP yang muncul pada metode *linking* tersebut. Dari hasil penelitian ini dapat dinyatakan bahwa (1) penyamaan skala dengan metode RS lebih akurat dari metode TRS, (2) penyamaan skala dengan metode RS memiliki keakuratan yang sama dengan metode KK, (3) penyamaan skala dengan metode TRS kurang akurat dibandingkan metode KK.

Perbedaan hasil deteksi *DIF* butir ke- j yang penyamaan skalanya menggunakan metode tegar rerata dan sigma (TRS), metode rerata dan sigma (RS) dan metode kurva karakteristik (KK) terletak pada nilai hasil penyamaan taraf sukar butir (b_F) dan daya beda butir (a_F) kelompok fokal dari ketiga metode tersebut, sedangkan nilai faktor kebetulan betul (c_F) memiliki nilai yang sama. Pendeteksian *DIF* pada butir ke- j ditentukan oleh uji statistik $\chi_j^2 = v_j' \Gamma_j^{-1} v_j$, $v_j = (\bar{a}_r - \bar{a}_n \quad \bar{b}_r - \bar{b}_n \quad \bar{c}_r - \bar{c}_n)$ matriks vektor perbedaan estimasi parameter butir ke- j antara kelompok referensi dan kelompok fokal dan Γ_j^{-1} invers dari jumlah matriks va-riansi kovariansi butir ke- j kelompok referensi dan fokal. Apabila perkalian ketiga matriks itu disederhanakan maka akan diperoleh suatu fungsi kuadratik. Variabel bebas dari fungsi kuadrat butir ke- j adalah taraf sukar butir dan daya beda butir ke- j kelompok fokal. Dengan demikian nilai fungsi kuadrat butir ke- j dipengaruhi oleh nilai taraf sukar butir dan daya beda butir dari kelompok fokal yang telah disamakan skalanya baik dengan menggunakan metode TRS, RS atau KK. Semua nilai taraf sukar butir (b_F) dan daya beda butir (a_F) yang memiliki nilai fungsi $f(a_F, b_F) - 7,8 > 0$ dinyatakan sebagai butir ke- j *DIF*.

Sebagai contoh butir ke-33 pada replikasi 58 terdeteksi *DIF* yang penyamaan skalanya dengan menggunakan metode KK, sedangkan butir ini tidak terdeteksi *DIF* yang penyamaan skalanya menggunakan metode TRS dan metode RS. Nilai hasil penyamaan taraf sukar butir, daya beda butir dan faktor kebetulan betul dengan metode KK adalah 0,906542, 0,166662 dan 0,157. Nilai hasil penyamaan taraf sukar butir, daya beda butir dan faktor kebetulan betul dengan metode TRS adalah 1,018950,

0,265916, dan 0,157. Nilai hasil penyamaan taraf sukar butir, daya beda butir dan faktor kebetulan betul dengan metode RS adalah 1,096527, 0,275345 dan 0.157. Nilai hasil penyamaan taraf sukar butir, daya beda butir kelompok focal dari ketiga metode memiliki nilai yang berbeda, sedangkan nilai faktor kebetulan betul memiliki nilai yang sama.



Gambar 2. Permukaan

Terdeteksinya *DIF* pada butir ke-33 yang penyamaan skalanya menggunakan metode KK disebabkan oleh nilai taraf sukar butir (a_F) dan daya beda butir (b_F) yang memiliki nilai $f(a_F, b_F) - 7,8 = 37,32 a_F^2 + 124,05 b_F^2 - 63,98 a_F - 114,27 b_F + 11,58 a_F b_F + 42,27 > 0$. Sementara nilai taraf sukar butir (a_F) dan daya beda butir (b_F) ke-33 yang penyamaan skalanya menggunakan metode TRS dan metode RS memiliki nilai $f(a_F, b_F) - 7,8 = 37,32 a_F^2 + 124,05 b_F^2 - 63,99 a_F - 114,27 b_F + 1,58 a_F b_F + 42,27 < 0$. Semua nilai taraf sukar butir (a_F) dan daya beda butir (b_F) ke- 33 yang merupakan hasil dari penyamaan skala pada pengulangan 58, apabila memiliki nilai fungsi $f(a_F, b_F) - 7,8 = 37,32 a_F^2 + 124,05 b_F^2 - 63,99 a_F - 114,28 b_F + 11,59 a_F b_F + 42,27 > 0$ maka dinyatakan bahwa butir tersebut mengandung *DIF*. Gambar fungsi ini dapat dilihat pada Gambar 2.

Simpulan

Keakuratan ketiga metode *linking* pada pendeteksian *DIF* melalui metode *Lord Chi-Square* dilihat dari banyaknya butir *FP*. Semakin sedikit butir *FP* yang muncul pada metode RS, TRS dan KK maka disimpulkan bahwa salah satu metode *linking* tersebut paling akurat. Dengan demikian, maka berdasarkan pengujian hipotesis dapat disimpulkan bahwa metode KK dan metode RS lebih akurat dari metode TRS, sedangkan metode KK memiliki keakuratan yang sama dengan metode RS.

Sampel butir dalam penelitian ini memuat 11 butir tidak *DIF* yang dideteksi dengan menggunakan metode *Likelihood Ratio Test*. Dengan kata lain, proposi butir *DIF* pada sampel butir sebanyak 44 %. Candell dan Drasgow (1988) menyatakan persentase banyaknya butir *DIF* akan menaikkan banyaknya butir *false positive (FP)*. Untuk itu perlu dilanjutkan dengan penelitian lanjutan mengenai keakuratan metode *linking* dengan memperhatikan proporsi jumlah butir *DIF*, jumlah butir jangkar, ukuran sampel pada model logistik tiga parameter (L3P), model logistik dua parameter (L2P) atau model logistik satu parameter (L1P).

Penelitian yang berhubungan dengan pendekteksian *DIF* melalui metode *Lord's Chi-Square*. Untuk mendeteksi suatu butir mengandung *DIF* atau tidak maka terlebih dahulu dilakukan penyamaan skala parameter butir kedua kelompok, kemudian dilanjutkan dengan deteksi *DIF*. Pada metode *Lord's Chi-Square* penyamaan skalanya dilakukan secara terpisah. *Penyamaan* skalanya menggunakan metode KK atau RS.

Tes prestasi belajar yang berskala luas seperti Ujian Nasional (UN) dan Seleksi Penerimaan Mahasiswa Baru (SPMB). Butir-butir di dalam perangkat tes ini harus diusahakan seminimal mungkin mengandung bias. Oleh karena itu butir-butir tersebut harus dianalisis *DIF*. Apabila butir-butir dalam perangkat tes dideteksi dengan menggunakan metode *Lord's Chi-Square* maka penyamaan skalanya menggunakan metode KK atau RS.

Hasil penelitian ini mengembangkan software penyamaan skala dengan metode RS dan TRS serta *software* deteksi *DIF* dengan metode *Lord Chi-Square Software* yang digunakan untuk penyamaan skala dinamakan

BLINK dan software yang digunakan untuk pendeteksian DIF dengan menggunakan metode *Lord Chi-Square* dinamakan LORDDIF.

Daftar Pustaka

- Allen, Mary J. dan Wendy M Yen, *Introduction to Measurement Theory*, California: Broke/California, 1979.
- Badrun, “*Perbandingan Berbagai Metode untuk Mendeteksi Bias Butir*”. Disertasi, Universitas Gajah Mada, 2004.
- Baker, Frank B. *Equate Computer Program*. Madison Wisconsin: Wisconsin University, 1995.
- Baker, Frank K. *The Basic of Item Response Theory*. Diambil pada tanggal 8 April 2005 <http://webs.ono.com/usr021/Jose.L.Melia/Papers/Baker2001.pdf>.
- Berk, Ronald A. *Handbook of Methods for Detecting Test Bias*. London : The Johns Hopkins University Press.
- Camili, Gregory dan A Shepard Lorrie. *Methods for Identifying Biased Test Items*. London: Sage Publications, 1994.
- Candell, Gregory L.dan Frits Drasgow. *An Iterative Procedure for Linking Metrics and Assensing Item Bias in Item Response Theory*. Journal Applied Psychological Measurement Vol 12 No 3 September 1988.
- Cohen, Allan S. dan Seock-Ho Kim. *Effect of Linking Methods on Detectoin of DIF*”. Journal of Educational Measurement, Vol. 29 No. 1 Musim Semi. 1992
- Cohen, Allan S. dan Seock-Ho Kim. *An Investigation of linking Methods Under the Graded Response Model*. Journal Applied Psychological Measurement. Vol 22 No 2 Juni 1998.
- Cohen, Allan S. dan Seock-Ho Kim. *Comparison of Linking and Concurenrent Calibration Under Item Rersponse Theory*. Journal Applied Psychological Measurement. Vol 22 No 2 Juni 1998.

- Cohen, Allan S. dan Seock-Ho Kim. "The Equating Under The Multiple-Choice Model". *Journal Psychological Measurement*, Vol 26 No 3, September 2002.
- Cohen, Allan S. dan Seock-Ho Kim. "*Detectin of Differential Item Functionong Under the Graded Response Model With The Likelihood Ratio Test*". *Journal Applied Psychological Measurement*. Vol 22 No 4 Desember 1998
- Crocker, Linda dan James Algina. "*Introduction To Classical and Modern Test Theory*". New York : cbs College Publishing, 1986.
- Daniel, Wayne W. "*Applied Nonparametric Statistics*". Bonston : PWS Kent Publishing Company, 1990.
- McDonald dan Roderick P.. "*Test Theory: A Unified Treatment*". London : Lawrence Erlbaum Associates, 1999.
- Divgi, D. R.. "*Minimum Chi-Square Method for Developing a Common Metric in Item Response Theory*". *Journal Applied Psychological Measurement*. Vol 9 no. 4. Desember 1985.
- Fidalgo, Angel M dan Gideon J Mellenbergh, Jose Muniz. "*Effects of Amount of DIF, Test Length, and Purification Type on Robustness and Power of Mantel-Haenszel*". Diambil pada tanggal 10 September 2003, dari <http://www.mpr-online.de>
- Gregory, Robert J. "*Psychological Testing. History, Principles and Applications*". ____: Third edition, 2000.
- Hambleton, Ronald K. "*Principles and Selected Applications of Item Response Theory*". New York : Macmillan, 1989.
- Hambleton, Ronald K, H Swaminathan, dan H Jane Rogers. "*Fundamentals of Item Response Theory*". London : Sage Publications, 1991.
- Hambleton, Ronald K dan H Swaminathan, "*Item Response Theory: Principles and Applications*". Boston : Kluwer.Nijhoff Publishing, 1990.

- Hulin, Charles L, Fritz Drasgow dan Charles K Parsons. *Item Response Theory : Application to Psychological Measurement*. Homewood, Illinois: Dow Jones-Irwin, 1993.
- Isaac, Stephen dan William B Michael. *Handbook In Research and Evaluation*. California: EDITS, 1995.
- Kolen, Michael dan Robert L Brennan. *Test Equating*. New York: Springer, 1995.
- Li, Hsin-Hung dan William Stout. 1995. *New Prosedure for Detecting of Crossing DIF*. University of Illinois at Urbana-Champaign online. Tanggal 17 Juni 2003 dari <http://www.stat.uiuc.edu/paper/Li941.pdf>.p 1.
- Linn, Robert L.. *Educational Measurement : Third Edition*. London: Collier Macmillan Publishers, 1989.
- Lord, Federic M. *Applications of Item Response Theiry to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates Publihers. 1980.
- MacIntosh, Randall dan Sabrina Hashim. Variance Estimatin for Converting MIMIC Model Parameters to IRT Parameters in DIF Analysis. *Journal Applied Psychological Measurement*. Vol 27 No 5 September 2003.
- Mattjik, Ahmad Ansori dan Made Sumertajaya,. *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*. Bogor : IPB Press, 2000.
- McLaughlin, Mary E dan Fritz Drasgrow,. *Lord's Chi-Square of Item Bias With Estimated and With Know Person Parameters*. *Journal Applied Psychological Measurement* Vol. 11 No. 2. June 1987.
- Mislevy, Robert J dan R Darrell Bock. *Bilog 3: Item Analysis and Test Scoring with Banary Logistic Models*. Mooresville: Scientific Software, 1990
- Naga, Dali S. *Pengantar Teori Sekor Pada Pengukuran* Jakarta: Besbtas, 1992.
- Naga, Dali S. *Karakteristik Butir Pada Alat Ukur Model Dikotomi*. *Jurnal Ilmiah Psikologi "ARKHE"* Th 3/no 4/ 1998.

- Naga, Dali S. Karakteristik *Butir Pengukuran Pada Teori Kalsik dan Teori Modern*. *Jurnal Ilmiah Psikologi*"ARKHE" Th 3/no 5/ 1998
- Naga, Dali S. *Estimasi Parameter Responden Pada Model L2P Melalui Kebolehjadian Maksimum dan Metode Newton-Raphson*. *Jurnal Ilmiah Psikologi*"ARKHE" Th 4/no 7/ 1999.
- Ogasawara, Haruhiko. *Standard Errors of Item Response Theory Equating/Linking by Response Fubction Mathods*. *Journal Applied Psychological Measurement*. Vol 25 No 1 Maret 2001.
- Popham, W James. *Modern Educational Measurement..* Carlifornia: Prentice-hall.Inc, Engewood Cliffs, 1981
- Susangko, Purwa dan Djemari Mardapi,. *Keberfungsian Butir Diferential Perangkat Tes Ebtanas Kimia Sekolah Menengah Umum di Jawa Tengah*. *Jurnal Penelitian dan Evaluasi*, tahun III. 2001.
- Stark, S. dan Oleksandr Chernyshenko, . *Detection of differential item/test functioning (DIF/DTF) Using IRT*. University of Illinois at Urbana Champaign online. Diambil pada tanggal 10 September 2003, dari <http://www.work.psych.uiuc.edu/irt/>.
- Swaminatahn, H dam Hambleton , Ronald. K, Stephen G Sireci et al. *Small Sample Estimation in Dichotomous Item Response Models : Effect of Prior Based on Judgmental Information on the Accuracy of Item Parameter Estimates*. *Journal Applied Psychological Measurement*. Vol 27 No 1 Januari 2003.
- Swediati, Nony, "*Equiating Tests Under the Generalized Partial Credit Model*", Dissertation, University of Massachusetts, 1997.
- Veerkamp, Wim J.J. dan Berger, Martijn P.F.. *Optimal Item Discrimination and Maximum Information for Logistic IRT Models*. *Journal Applied Psychological Measurement*. Vol 23 No 1 Maret 1999.
- Wang, Wang-Chung dan Ya-Li Yeh. *Effects of Anchor Item Methods on Differential Item Functioning Detection With the Likelihood Ratio*.