

**DETECTING CROSSING DIFFERENTIAL ITEM FUNCTIONING (CDIF):
BASED ON ITEM RESPONSE THEORY**

Effendi

Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih KM. 32 Indralaya, Ogan Ilir
effendi_unsri@yahoo.com

DOI: doi.org/10.21009/JEP.022.03

Abstract

The objective of this study was to find out the empirical data of the index of crossing Differential Item Functioning (DIF) and the difference of sensitivity among methods in detecting crossing DIF according to the Item Response Theory. The study used quasi experimental methods with 1 x 3 designs. The independent variables were Raju's area measure, Lord's chi-square, likelihood ratio test, and sample size, while the dependent variable was index of crossing DIF. Methods for detecting uniform DIF had been well developed, where as those for identifying the presence of crossing DIF had not. The three methods compared were Raju's area measure, Lord's chi-square, and likelihood ratio test. Factors manipulated were sample size, ability differences between two groups, percentage of DIF, and the examinee's response. The results indicated that all methods were good for the detection of crossing DIF, but Raju's area measure was the most sensitive among those methods compared. Besides, there was no DIF caused by gender differences among the examinees with great ability.

Keywords: *crossing DIF, sample size, Raju's area measure, Lord's chi-square, likelihood ratio test*

PENDETEKSIAN *CROSSING DIFFERENTIAL ITEM FUNCTIONING* (CDIF): BERDASARKAN TEORI RESPONSI BUTIR

Effendi

Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih KM. 32 Indralaya, Ogan Ilir
effendi_unsri@yahoo.com

Abstrak

Tujuan utama dari penelitian ini adalah untuk mendapatkan data empirik tentang indeks dari *crossing differential item functioning* (CDIF) dan perbedaan sensitifitas antara metode pendeteksian CDIF berdasarkan teori responsi butir. Penelitian ini menggunakan metode kuasi eksperimen dengan desain 1 x 3. Variabel bebas adalah Luasan menurut Raju, Khi-kuadrat menurut Lord, tes rasio kebolehjadian, serta ukuran sampel, sedangkan variabel terikatnya adalah indeks CDIF. Metode untuk pendeteksian *uniform DIF* telah banyak dikembangkan, sedangkan untuk pendeteksian CDIF belum banyak dikembangkan. Ketiga metode yang dibandingkan adalah Luasan menurut Raju, Khi-kuadrat menurut Lord, dan Tes Rasio Kebolehjadian. Faktor yang dimanipulasi adalah ukuran sampel, perbedaan kemampuan responden dari dua kelompok, persentase CDIF, serta hasil tes responden. Hasil penelitian menunjukkan bahwa ketiga metode dapat digunakan untuk pendeteksian CDIF, tetapi Metode Luasan menurut Raju sangat sensitif dalam mendeteksi CDIF dibandingkan dengan kedua metode yang lain. Disamping itu tidak ada DIF yang disebabkan perbedaan gender pada tingkat kemampuan tinggi.

Kata Kunci: CDIF, ukuran sampel, luasan menurut Raju, Khi-kuadrat menurut Lord, tes rasio kebolehjadian

PENDAHULUAN

Evaluasi bertujuan untuk mengetahui hasil belajar peserta didik pada tingkat satuan pendidikan tertentu, dengan menggunakan ukuran yang ditetapkan pada akhir masa pendidikannya. Pemerintah melalui Kementerian Pendidikan Nasional mengeluarkan Permen Nomor 19 Tahun 2005 pasal 63 ayat (1) tentang penilaian oleh pemerintah dalam bentuk Ujian Nasional (UN) yang dilakukan oleh Badan Standar Nasional Pendidikan (BNSP). Kenyataan di lapangan, terkadang skor pada hasil ujian tidak memberikan informasi yang benar tentang peserta ujian. Mungkin saja informasi itu tidak menjangkau sampai ke besaran atau dimensi yang hendak diukur oleh ujian itu. Mungkin pula, hasil ujian itu tercampur dengan besaran atau dimensi lain yang tidak dimaksudkan untuk diukur oleh ujian tersebut sehingga hasil ujian menjadi rancu atau terjadi ketimpangan skor (Naga, 1992: 435).

Banyak hal yang menyebabkan terjadinya ketimpangan skor di dalam pengukuran pendidikan, namun secara garis besar, penyebab itu dapat dikelompokkan menjadi dua, yaitu; (1) masing-masing peserta tes dan penyelenggara tes, dan (2) perangkat tes beserta butir-butirnya. Terkait dengan hal ini Naga (1992: 437) menjelaskan bahwa ketimpangan itu dapat terjadi pada peserta tes atau butir tes, atau secara individu atau secara kelompok. Dua pasang alternatif ini dapat membentuk empat jenis kombinasi ketimpangan skor, yaitu; (1) ketimpangan skor pada peserta secara individu atau seorang demi seorang, (2) ketimpangan skor pada peserta secara kelompok, (3) ketimpangan skor pada butir secara individu atau satu demi satu, dan (4) ketimpangan skor pada butir secara kelompok atau pada tes. Ketimpangan skor pada kombinasi ketiga dan keempat disebut bias butir. Hambleton dkk (1991: 109) mengemukakan pengertian bias yaitu individu-individu yang memiliki kemampuan sama dari kelompok berbeda mempunyai probabilitas yang tidak sama untuk dapat menjawab dengan betul butir tersebut. Susongko dan Mardapi seperti dikutip oleh Rahayu (2008: 3) langkah-langkah yang dilakukan untuk mendeteksi suatu butir bias atau tidak yaitu menemukan butir yang berpotensi bias dengan menggunakan analisis statistika. Selanjutnya butir yang berpotensi mengandung bias kemudian dianalisis dengan analisis logik, misalnya mengapa butir ini relatif sulit untuk satu kelompok. Apakah suatu butir relatif lebih sulit untuk suatu kelompok dan kesulitan tersebut tidak relevan terhadap konstruk tes maka butir tersebut dikatakan bias.

Ahli lain menggunakan istilah *differential item functioning (DIF)* untuk butir yang berpotensi mengandung bias. Angoff (1993: 4-5) menyatakan bahwa *DIF* adalah keberfungsian yang berbeda dari suatu butir uji tes. Hambleton dkk (1991: 110) menggunakan istilah definisi *DIF* secara operasional yang dihubungkan dengan kurva karakteristik butir, yaitu jika butir menunjukkan *DIF* jika kurva karakteristik butir pada subkelompok berbeda tidak berhimpit dan sebaliknya suatu butir tidak menunjukkan *DIF* jika kurva karakteristik butir dari subkelompok yang berbeda adalah berhimpit. Hal yang sama dikemukakan oleh Lord (1980: 213) yaitu suatu butir menunjukkan *DIF* jika dua kurva karakteristik butir dari dua kelompok berbeda tidak berhimpit.

Ada dua jenis *DIF*, yaitu *DIF* seragam (*Uniform DIF*) dan *Crossing DIF (CDIF)*. Finch dan French (2007: 556) menyatakan bahwa suatu butir menunjukkan *DIF* seragam (*Uniform DIF*), jika kurva karakteristik butir suatu kelompok lebih tinggi dari kelompok yang lain. Sedangkan *CDIF* jika kedua kurva karakteristik butir berpotongan atau kurva karakteristik butir satu kelompok pada selang kemampuan tertentu lebih tinggi dari kelompok yang lain pada selang kemampuan yang lain, menguntungkan kelompok yang lain pula.

Untuk pendeteksian keberadaan *DIF* pada butir tes, sebuah populasi dibagi menjadi dua kelompok, yaitu kelompok fokal dan kelompok referensi. Kelompok

fokal merupakan kelompok yang diselidiki apakah ada butir yang mengandung *DIF* pada kelompok tersebut. Kelompok referensi merupakan kelompok pembanding. Kedua kelompok diambil dari populasi yang berbeda dan mengerjakan butir pada perangkat tes yang sama. Pembagian kelompok didasarkan pada jenis kelamin, wilayah (kota dan daerah), ras (kulit hitam atau kulit putih), dan etnis. Pada penelitian ini pembagian kelompok populasi berdasarkan jenis kelamin. Salah satu kelompok yang paling banyak digunakan dalam pendeteksian adanya butir *DIF* adalah jenis kelamin. Halpern (1997: 1091) menjelaskan adanya perbedaan antara sex atau jenis kelamin. Jenis kelamin menunjuk pada perbedaan biologis dan fisik antara wanita dengan pria, sedangkan jenis kelamin menunjuk pada ciri-ciri psikis dan tingkah laku serta memperhatikan budaya. Peranan jenis kelamin menunjuk pada tingkah laku yang diharapkan dari pria dan wanita dikaitkan dengan perbedaan biologis dan reproduksinya. Peranan jenis kelamin menunjuk pada tingkat kemampuan seseorang mengadopsi tingkah laku jenis kelamin yang khusus dan cocok yang dianggap berasal dari budayanya.

Secara formal pendeteksian butir yang mengandung *DIF* dilakukan dengan menggunakan analisis statistika. Ada beberapa metode yang dapat digunakan untuk pendeteksian butir *DIF* berdasarkan teori responsi butir, yaitu; Hambleton dkk (1985: 256) menjelaskan bahwa ada satu metode untuk mendeteksi *DIF* dengan pendekatan IRT, yaitu metode kurva karakteristik butir. Sementara itu Naga (1992: 258) menjelaskan ada dua metode untuk mendeteksi *DIF*, yaitu metode kurva karakteristik butir dan metode luas antara lengkungan. Kim dan Cohen (1991: 456) menambahkan bahwa paling tidak ada tiga metode untuk mendeteksi *DIF* yaitu: (1) metode luasan menurut Raju, (2) *chi-square* menurut Lord, dan (3) tes rasio kebolehjadian. Pada penelitian ini digunakan tiga metode pendeteksian *DIF* yaitu: (1) metode luasan menurut Raju, (2) *chi-square* menurut Lord, dan (3) tes rasio kebolehjadian.

Metode-metode tersebut dapat digunakan untuk pendeteksian *DIF* pada tes penilaian pendidikan yang dilakukan oleh pemerintah, seperti Ujian Nasional. Penilaian hasil belajar berbentuk Ujian Nasional bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi dilakukan dalam bentuk Ujian Nasional. Ujian Nasional dilakukan secara obyektif, berkeadilan, dan akuntabel (Peraturan Pemerintah No. 19 Tahun 2005). Idealnya tidak ada kesalahan dalam pengukuran yang dilakukan oleh peserta uji tes, pelaksanaan tes, dan juga tidak ada kesalahan pengukuran yang disebabkan oleh butir tes. Instrumen UN yang digunakan untuk mengukur seharusnya memiliki validitas dan reliabilitas yang tinggi dan adil. Artinya tidak ada orang atau kelompok orang tertentu yang dirugikan dengan adanya butir tes yang tidak adil itu. Untuk mengatasi agar butir yang bias

terdeteksi sedini mungkin, maka diperlukan metode pendeteksian *uniform DIF* dan *CDIF* yang sensitif.

Hasil survei menunjukkan bahwa pada pelaksanaan UN yang dilakukan oleh pemerintah, sangat banyak terjadi ketimpangan skor, terutama skor peserta yang berasal dari daerah dengan yang berasal dari kota, kemungkinan terjadi juga pada peserta kelompok pria dan kelompok wanita. Hal inilah yang menarik untuk dilakukan penelitian dengan bertitik tolak dari suatu pertanyaan mendasar. Apakah ada metode yang paling sensitif dalam pendeteksian *DIF* khususnya *CDIF*, dan dalam penelitian ini apakah ada perbedaan sensitifitas antara metode luasan menurut Raju, metode *chi-square* menurut Lord, dan tes rasio kebolehdjian dalam pendeteksian *CDIF*. Penelitian ini hanya memfokuskan pada *CDIF* yang disebabkan oleh perbedaan jenis kelamin, serta untuk mata pelajaran kimia, sebab mata pelajaran kimia merupakan salah satu mata pelajaran yang diujikan secara nasional dalam UN.

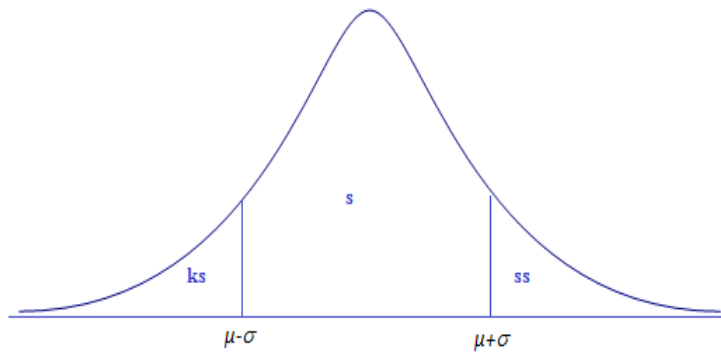
METODE PENELITIAN

Data penelitian berupa nilai indeks *CDIF* yang diperoleh melalui metode pendeteksian *CDIF*, yaitu metode *chi-square* menurut Lord, luasan menurut Raju, dan tes rasio kebolehdjian menggunakan responden jawaban siswa pada Ujian Nasional tahun pelajaran 2007/2008 mata pelajaran kimia paket utama 11 kimia (D14) SMA/MA Program Studi IPA di SMA Negeri dalam wilayah DKI Jakarta. Sebelum melangkah menganalisis data hasil penelitian, terlebih dahulu dilakukan uji kecocokan model yakni menguji apakah butir tes cocok dengan model logistik tiga parameter (L3P).

Untuk persiapan uji hipotesis statistik dilakukan terlebih dahulu uji persyaratan analisis yaitu uji unidimensi, independensi lokal, dan uji kecocokan model. Pengujian unidimensi dan independensi lokal digunakan program excel dan uji kecocokan model digunakan program ASCAL. Setelah persyaratan model dan persyaratan analisis dilakukan maka baru dilakukan pengujian hipotesis.

Analisis data penelitian dilakukan dengan menggunakan metode analisis statistika uji-z. Selanjutnya dilakukan analisis tingkat sensitifitas dengan menggunakan kriteria acuan normal, yaitu untuk melihat seberapa besar sensitifitas masing-masing metode digunakan penilaian yang berdasarkan kriteria acuan normal, yakni dibuat rentang distribusi normal dengan mengambil batas atas rerata gabungan dua variabel yang independen ditambahkan dengan simpangan baku gabungan atau $\mu+\sigma$, dan batas bawah rerata gabungan dua variabel yang independen dikurangi dengan simpangan baku gabungan atau $\mu-\sigma$, kemudian dibagi sesuai kriteria yang diinginkan, yakni: Jika nilai rerata kelompok lebih besar daripada nilai $\mu+\sigma$ dikatakan sangat sensitif, nilai rerata kelompok di antara nilai $\mu+\sigma$ dan $\mu-\sigma$

dikatakan sensitif, dan nilai rerata kelompok lebih kecil dari nilai $\mu - \sigma$ dikatakan kurang sensitif.



Gambar 1. Kriteria Acuan Normal

Keterangan : SS (sangat sensitif), S (sensitif), dan KS (kurang sensitif)

HASIL PENELITIAN

Hasil Pendeteksian *CDIF*

Dari hasil analisis pendeteksian total butir yang mengandung *CDIF* dengan menggunakan tiga metode pendeteksian, yaitu; metode luasan menurut Raju, metode *chi-square* menurut Lord, dan metode tes rasio kebolehjadian, dirangkum pada tabel 1 berikut :

Tabel 1. Distribusi Butir Crossing DIF

No.butir	Luasan Raju		<i>Chi-square</i> Lord		Tes rasio kebolehjadian	
	F	P	F	P	F	P
04	48	96%	46	92%	24	48%
09	37	74%	9	18%	4	8%
13	46	92%	23	46%	9	18%
36	49	98%	43	86%	32	64%
Rata-rata	45	90%	30,25	60,5%	17,25	34,5%

Keterangan: f adalah frekuensi, p=persentase

Dengan mengamati tabel 1 maka dapat diketahui bahwa ada perbedaan jumlah yang terdeteksi oleh masing-masing metode pendeteksian *CDIF*. Metode luasan menurut Raju hampir dapat mendeteksi semua butir yang mengandung *CDIF*,

yakni butir nomor 36, nomor 4, dan butir nomor 13 terdeteksi sempurna yaitu 98%, 96%, dan 92% terdeteksi, hanya butir nomor 9 yang 74% dengan rata-rata muncul 45 kali atau 90% dari total pengulangan. Hal ini menunjukkan bahwa metode ini sangat baik untuk mendeteksi *CDIF*. Metode *chi-square* menurut Lord hanya butir nomor 4 terdeteksi sempurna yakni 92% terdeteksi, dan butir nomor 36 terdeteksi hampir sempurna yaitu 86% terdeteksi, sedangkan butir nomor 9 dan nomor 13 di bawah 50% terdeteksi, rata-rata muncul 30,25 atau 60,5%. Hal ini menunjukkan bahwa metode *chi-square* menurut Lord dapat dikatakan baik untuk mendeteksi *CDIF*. Sedangkan metode tes rasio kebolehan dengan rata-rata muncul hanya 17,25 atau 34,5% dari total pengulangan. Hanya butir nomor 36 yang terdeteksi di atas 50% yakni 64% sedangkan butir yang lain di bawah 50% terdeteksi. Bahkan butir nomor 9 hanya 8% terdeteksi dari 50 kali pengulangan. Hal ini menunjukkan bahwa metode tes rasio kebolehan kurang baik jika digunakan untuk mendeteksi *CDIF*.

Pengujian Hipotesis

Hasil pengujian hipotesis mengenai komparasi sensitifitas antara metode luasan menurut Raju dengan metode tes rasio kebolehan, metode *chi-square* menurut Lord dengan metode tes rasio kebolehan, dan metode luasan menurut Raju dengan metode *chi-square* menurut Lord adalah:

1. Uji hipotesis 1 adalah melihat perbandingan antara metode luasan menurut Raju dengan metode tes rasio kebolehan dengan kriteria tolak H_0 jika $Z_{hitung} > Z_{tab}$ pada $\alpha = 0,05$ dan hasil perhitungan sebagai berikut: $Z_{hitung} > Z_{tab}$ yaitu: $(5,7246 > 1,655)$ sehingga H_0 ditolak, yang berarti metode luasan menurut Raju lebih sensitif dalam pendeteksian *crossing DIF* daripada metode tes rasio kebolehan.
2. Uji hipotesis 2 adalah melihat perbandingan antara metode *chi-square* menurut Lord dengan metode tes rasio kebolehan dengan kriteria tolak H_0 jika $Z_{hitung} > Z_{tab}$ pada $\alpha = 0,05$ dan hasil perhitungan menunjukkan $Z_{hitung} > Z_{tab}$ yaitu: $(2,603 > 1,655)$ sehingga H_0 ditolak atau dengan kata lain metode *chi-square* menurut Lord lebih sensitif dalam pendeteksian *crossing DIF* daripada metode tes rasio kebolehan.
3. Uji hipotesis 3 adalah melihat perbandingan antara metode luasan Raju dengan metode *chi-square* menurut Lord dengan kriteria tolak H_0 jika $Z_{hitung} > Z_{tab}$ pada $\alpha = 0,05$ dan hasil perhitungan menunjukkan $Z_{hitung} > Z_{tab}$ yaitu: $(3,418 > 1,655)$ sehingga H_0 ditolak yang berarti metode luasan menurut Raju lebih sensitif dalam pendeteksian *crossing DIF* daripada metode *chi-square* menurut Lord.

Uji Tingkat Sensitifitas

Selanjutnya untuk uji tingkat sensitifitas dilakukan dengan melihat kriteria acuan normal, sebagai berikut: pertama dicari dulu nilai rerata gabungan (\bar{x}) dan standar deviasi gabungan s_p didapatkan nilai sebagai berikut:

$$\bar{x} - s_p = 2,5 - 0,573 = 1,927$$

$$\bar{x} + s_p = 2,5 + 0,573 = 3,073$$

Dengan mencermati nilai di atas maka dapat disimpulkan:

- metode luasan Raju dengan rerata sebesar 3,62 artinya berada pada posisi $\bar{x}_1 > (\bar{x} + s_p)$ wilayah sangat sensitif,
- metode *chi-square* menurut Lord dengan rerata sebesar 2,48 artinya berada pada posisi $(\bar{x} - s_p) > \bar{x}_2 > (\bar{x} + s_p)$ wilayah sensitif,
- metode tes rasio kebolehjadian dengan rerata sebesar 1,40 $\bar{x}_3 < (\bar{x} + s_p)$ wilayah kurang sensitif.

PEMBAHASAN

Total butir yang mengandung *DIF* pada tes UN Kimia SMA/MA tahun 2008 se-DKI Jakarta cukup besar, yaitu 16 butir yang mengandung *DIF*, yakni butir nomor 1, 2, 4, 6, 9, 10, 13, 16, 19, 20, 21, 22, 26, 27, 28, dan 36, atau 59,26% dari jumlah butir yang dideteksi *DIF*-nya sebanyak 27 butir, Sedangkan butir yang tidak cocok model logistik tiga parameter (L3P) sebanyak 13 butir, sehingga total butir yang perlu diperhatikan sebanyak 29 butir dari 40 butir yang diujikan secara nasional atau 72,50%. Hasil ini cukup memprihatinkan karena butir yang dideteksi adalah butir UN. Oleh karena itu perlu dicermati masing-masing butir yang mengandung *DIF* agar dapat ditindaklanjuti oleh pengembang tes UN.

Ada perbedaan jumlah butir *DIF* yang terdeteksi oleh masing-masing metode pendeteksian *DIF*, yaitu maksimum butir yang dapat dideteksi oleh metode luasan menurut Raju sebanyak 8 butir yang terdeteksi mengandung *DIF* atau 26,63%, metode *chi-square* menurut Lord sebanyak 7 butir yang terdeteksi mengandung *DIF* atau sekitar 25,93%, dan metode tes rasio kebolehjadian sebanyak 12 butir yang terdeteksi mengandung *DIF* atau 44,44%. Hal senada disampaikan oleh Kartowagiran (2005: 156) menyatakan bahwa metode pendeteksian *DIF* yang paling sensitif adalah metode tes rasio kebolehjadian, dengan jumlah butir yang terdeteksi mengandung *DIF* sebanyak 9 butir dari 10 butir yang terdeteksi mengandung *DIF* (32,14%). Karto-wagiran memfokuskan penelitiannya pada; (1) butir-butir tes Ujian Akhir Nasional (UAN) matematika SMP se-DIY tahun 2003, yang secara signifikan

mengandung *DIF* bila dideteksi dengan metode kurva karakteristik butir, metode luasan menurut Raju, metode *chi-square* menurut Lord, dan metode tes rasio kebolehjadian, (2) ada atau tidak ada perbedaan sensitivitas dari keempat metode itu dalam mendeteksi bias butir, (3) metode deteksi bias yang paling sensitif untuk tes UAN matematika SMP. Lebih lanjut Kartowagiran menegaskan bahwa ada 10 butir mengandung *DIF* dari 28 butir yang dideteksi *DIF*-nya, atau 35,71% butir-butir tes UAN matematika SMP tahun 2003 se-DIY yang terdeteksi mengandung *DIF*, hasil ini sangat memprihatinkan. Di samping itu siswa yang memiliki kemampuan tinggi ternyata juga tidak ada *DIF* yang disebabkan oleh jenis kelamin.

Rahayu (2008: 163) melakukan penelitian yang sama tentang *DIF*. Hasil penelitiannya menunjukkan bahwa metode tes rasio kebolehjadian dapat mendeteksi sebanyak 21 butir yang mengandung *DIF* dari 36 butir yang dideteksi *DIF* nya. Metode *chi-square* menurut Lord dapat mendeteksi 6 butir yang mengandung *DIF* dari 14 butir tidak mengandung *DIF* menurut metode tes rasio kebolehjadian. Dengan demikian dapat dikatakan bahwa ada 27 butir yang terdeteksi mengandung *DIF* dari 40 butir yang diujikan, atau 67,50% butir-butir pada Ujian Nasional matematika SMP paket 3 untuk wilayah Jakarta Timur dan Jakarta Barat yang dilaksanakan pada tahun 2004 positif mengandung *DIF*. Kartowagiran (2008) yang menyatakan bahwa metode yang paling sensitif dalam pendeteksian *DIF* adalah metode tes rasio kebolehjadian. Hal ini ditunjukkan dengan adanya butir yang terdeteksi mengandung *DIF*, yaitu dari 10 butir yang terdeteksi mengandung *DIF* 9 butir terdeteksi oleh metode tes rasio kebolehjadian. Jika dicermati lebih lanjut hasil penelitian Kartowagiran ternyata dari 9 butir yang terdeteksi mengandung *DIF* hanya 1 butir yang merupakan butir *CDIF* dan 8 butir yang lainnya adalah butir *DIF* seragam atau *uniform DIF*. Hal yang sama disampaikan oleh Rahayu (2010: 29) yang menunjukkan bahwa dari 21 butir yang terdeteksi dengan metode tes rasio kebolehjadian hanya 4 butir yang merupakan *CDIF* sisanya 17 butir adalah *uniform DIF*.

Hasil penelitian ini menunjukkan bias butir yang disebabkan jenis kelamin itu tidak selalu benar. Terbukti bahwa keempat hipotesis ini wajar karena dua hal, (1) cara penjabaran konsep bias oleh masing-masing metode, dan (2) cara penyetaraan yang berbeda-beda untuk masing-masing metode. Untuk indeks luasan menurut Raju, bias dijabarkan sebagai luasnya bidang yang dibatasi oleh dua kurva karakteristik butir yang dihasilkan oleh dua kelompok yang dikaji. Dalam metode *chi-square* menurut Lord, bias dijabarkan sebagai perbedaan parameter butir antara hasil responsi dari kelompok fokus dan hasil responsi dari kelompok referensi yang memiliki kemampuan sama. Sedangkan metode tes rasio kebolehjadian. Bias dijabarkan sebagai perbandingan antara kebolehjadian maksimum model kecil dengan kebolehjadian maksimum model besar, atau dapat juga diartikan sebagai perbandingan antara ketelitian model kecil (*compact model*) dengan parameter

tertentu dan ketelitian model besar (*augment model*) dengan parameter tertentu. Finch dan French (2007: 565) meneliti tentang pendeteksian *CDIF* dengan membandingkan 4 metode, yaitu *simultaneous item bias test (SIBTEST)*, *logistic regression*, *item response theory likelihood ratio test*, and *confirmatory factor analysis*. Hasil penelitiannya menunjukkan bahwa ke 4 metode pendeteksian *crossing DIF* baik, akan tetapi metode *SIBTES* sangat baik daripada ketiga metode yang lain. Penelitian yang sama menggunakan metode *IRTLR* untuk pendeteksian *CDIF* seperti yang dilakukan oleh Thissen (1993).

Pemerintah melalui Kementerian Pendidikan Nasional dengan berbagai peraturan menteri telah mengeluarkan standar penilaian. Tujuan dikeluarkannya standar penilaian adalah agar proses penilaian benar-benar berfungsi sebagaimana mestinya, terutama dalam pencapaian kompetensi dan kompetensi dasar yang harus dicapai oleh siswa, penilaian hasil belajar yang dilakukan oleh pemerintah dalam UN sesuai Peraturan Pemerintah Nomor 19 Tahun 2005. Sesuai dengan hasil penelitian, maka butir-butir yang ada dalam perangkat uji tes seperti UN harus diusahakan seminimal mungkin mengandung *DIF*, oleh karena itu butir-butir tes tersebut harus dianalisis *DIF*-nya terutama *CDIF* sebelum diajukan di lapangan.

Secara teoretis dapat dikembangkan metode pendeteksian *CDIF* melalui metode luasan menurut Raju. Untuk mendeteksi butir mengandung *CDIF* atau tidak maka terlebih dahulu dilakukan penyamaan skala parameter butir kedua kelompok, sebab analisis parameter butir dilakukan secara terpisah atau sendiri-sendiri. Penyamaan skalanya dilakukan dengan menggunakan metode kurva karakteristik atau metode rerata dan sigma (Rahayu, 2010: 32).

SIMPULAN

Melalui serangkaian kegiatan pengumpulan data dan analisis data dapat dirumuskan beberapa simpulan hasil penelitian yaitu: **pertama**, berdasarkan hasil pengujian hipotesis yang diajukan dalam penelitian ini dapat dibuktikan bahwa (1) metode luasan menurut Raju lebih sensitif dalam pendeteksian *CDIF* daripada metode tes rasio kebolehjadian, (2) metode *chi-square* menurut Lord lebih sensitif dalam pendeteksian *CDIF* daripada metode tes rasio kebolehjadian, dan (3) metode luasan menurut Raju lebih sensitif dalam pendeteksian *CDIF* daripada metode *chi-square* menurut Lord, **kedua**, berdasarkan uji tingkat sensitifitas, maka diperoleh gambaran tentang tingkat sensitifitas masing-masing metode sebagai berikut (1) metode luasan menurut Raju sangat sensitif dalam mendeteksi *CDIF*, (2) metode *chi-square* menurut Lord sensitif dalam mendeteksi *CDIF*, dan (3) metode tes rasio kebolehjadian kurang sensitif dalam mendeteksi *CDIF*, **ketiga**, berdasarkan kurva karakteristik butir yang mengandung *DIF* dapat disimpulkan bahwa siswa dengan tingkat kemampuan tinggi yang berasal dari dua kelompok yang berbeda tetapi

memiliki tingkat kemampuan yang sama ternyata memiliki probabilitas menjawab betul yang sama pada suatu butir, atau dengan kata lain tidak ada *DIF* yang disebabkan oleh jenis kelamin.

DAFTAR PUSTAKA

- Angoff, William H. (1982). *Use of Difficulty and Discrimination Indices for Detecting Item Bias in Handbook of Methods for Detecting Test Bias*, ed. Ronard K. Beck. Baltimore: The Johns Hopkins University Press.
- Cohen, Allan S., dan Seock-Ho Kim. (1995). "A Comparison of Lord's chi square, Raju's area measures, and the Likelihood ratio test on detection of differential item functioning." *Journal of Applied Measurement in Education*. Vol. 4.
- Finch Holmes W. Dan Brian F., French. (2007). "Detection of Crossing Differential Item Functioning; A Comparison of Four Methods." *Journal Education and Psycho-logical Measurement*, Vol. 67 (4).
- Hambleton, Ronald K., H. Swaminathan, dan H. Jane Rogers. (1991). *Fundamentals of Item Response Theory*. London: SAGE Publication, Inc.
- Halpern, Diana F. (1997). "Sex Differences in Intelligence Implication for Education". *American Psychologist*, Vol. 52.
- Kartowagiran, Badrun. (2005). "Perbandingan Berbagai Metode untuk Mendeteksi Bias Butir." *Disertasi*, Universitas Gadjah Mada.
- Lord, Frederic M. (1980). *Aplication of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates Publisher.
- Naga, Dali S. (1992). *Pengantar Teori Skor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Rahayu, Wardani, (2008). "Pengaruh Metode *Linking* terhadap Banyak *Butir False Positive* pada Pendeteksian DIF menurut Teori Respon Butir." *Disertasi*, Jakarta: Program Pascasarjana UNJ.

_____. "Metode *Linking* dan Butir *False Positive* pada Pendekteksian *DIF* Berdasarkan Teori Responsi Butir." *Jurnal Penelitian dan Evaluasi Pendidikan*, Nomor 1 Tahun 2010.

Thissen, David, Lynne Steinberg, dan Howard Wainer. (1993). *Detection of Differential Item Functioning Using The Parameters of Item Response Models*, Di dalam *Differential Item Functioning* ed. Paul W. Halland dan Howard Wainer, Educational Testing Service. New Jersey: Lawrence Erlbaum Associates, Publisher.

Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan.