

# Pendekatan Regresi Logistik dalam Pendeteksian *DIF*

Oleh:

Kana Hidayati

Jurusan Pendidikan Matematika FMIPA UNY

## ABSTRAK

Kegiatan pengukuran dalam bidang pendidikan sangat memerlukan suatu alat ukur yang mengukur sesuai dengan apa yang hendak diukur dan tidak terpengaruh faktor luar yang tidak diukur. Dalam suatu alat ukur yang berupa tes, apabila terjadi keadaan dimana terdapat pengaruh dari faktor lain yang tidak diukur maka biasanya terjadi apa yang disebut dengan bias dari suatu butir tes atau dikenal dengan istilah keberfungsian butir diferensial (*Differential Item Functioning/DIF*).

Keberfungsian butir diferensial (*Differential Item Functioning/DIF*) merupakan kesalahan sistematis butir-butir tes yang dapat dideteksi lewat berbagai metode/prosedur. Metode/prosedur pendeteksian bias butir sangat berguna karena fokus terbesarnya adalah penelitian dengan cermat terhadap butir-butir tes secara individual dan menguji kembali butir-butir tes tersebut apakah berfungsi seperti yang diharapkan. Salah satu metode yang dapat digunakan dalam kegiatan pendeteksian *DIF* ini adalah pendekatan regresi logistik.

Kata kunci: Regresi logistik, *DIF*.

## A. Pendahuluan

Kegiatan pengukuran dalam bidang pendidikan sangat memerlukan suatu alat ukur yang valid dan reliabel. Hal ini dimaksudkan agar hasil pengukuran sesuai dengan apa yang hendak diukur dan tidak terpengaruh faktor luar yang tidak diukur. Untuk memperoleh suatu alat ukur yang demikian, khususnya yang berupa tes, maka perangkat tes yang disusun harus benar-benar dapat mengukur apa yang seharusnya diukur, memberikan hasil pengukuran yang dapat dipercaya, dan mencerminkan kemampuan siswa yang sebenarnya. Artinya, perbedaan skor yang diperoleh seorang siswa dengan siswa lainnya semata-mata hanya karena perbedaan kemampuan diantara mereka, bukan disebabkan oleh faktor lain, seperti adanya bias pada butir tes yang mereka kerjakan. Oleh karena itu, salah satu hal yang perlu diperhatikan adalah ada tidaknya bias pada butir-butir tes yang disusun. Apabila terjadi keadaan dimana terdapat pengaruh dari faktor yang tidak diukur maka biasanya terjadi apa yang disebut dengan bias dari suatu butir tes atau dikenal dengan istilah keberfungsian butir diferensial (*Differential Item Functioning/DIF*).

Terdeteksinya *DIF* pada suatu butir tes akan menyebabkan perangkat tes tersebut menjadi sesat karena bersifat tidak objektif, tidak adil, dan memihak kelompok tertentu sehingga dua siswa dari kelompok yang berbeda dengan kemampuan yang sama

---

Dipresentasikan dalam Seminar Nasional MIPA 2006 dengan tema "**Penelitian, Pendidikan, dan Penerapan MIPA serta Peranannya dalam Peningkatan Keprofesionalan Pendidik dan Tenaga Kependidikan**" yang diselenggarakan oleh Fakultas MIPA UNY, Yogyakarta pada tanggal 1 Agustus 2006

mempunyai peluang menjawab benar yang berbeda. Hal ini berakibat nilai yang diperoleh pun juga berbeda. Kondisi seperti ini merupakan permasalahan tersendiri yang berkaitan dengan kemampuan perangkat tes mengukur kemampuan siswa yang sebenarnya. Perangkat tes tersebut menjadi tidak baik kualitasnya karena mengandung sifat bias pada butir-butirnya. Oleh karena itu, pendeteksian terhadap ada tidaknya sifat bias pada butir-butir soal suatu perangkat tes sangat perlu untuk dilakukan agar ketidakadilan atau kerugian pada kelompok tertentu dapat dihindarkan dan kemampuan siswa terukur secara objektif.

Saat ini, penyusunan butir-butir tes pada umumnya telah dilakukan berdasarkan penelaahan butir tes secara kuantitatif yakni meliputi telaah materi, konstruksi, dan bahasa, baik oleh guru maupun ahli pendidikan lainnya. Secara teoritis perangkat tes tersebut dapat dikatakan telah baik dan dapat dipercaya. Namun demikian, secara empiris suatu perangkat tes yang baik hendaknya diuji secara kuantitatif. Nampaknya pengujian ini belum sepenuhnya dilakukan secara optimal. Keadaan yang demikian sangat memungkinkan terjadinya beberapa kelemahan pada perangkat tes yang digunakan.

Bias pada suatu butir tes yang dikenal sebagai *DIF* (*Differential Item Functioning*) atau keberfungsian butir diferensial dapat dideteksi dengan berbagai teknik atau metode pendeteksian *DIF*. Salah satu di antaranya yang cukup sering direkomendasikan oleh beberapa ahli daripada metode lainnya dalam pendekatan regresi logistik (Camilli & Shepard, 1994).

## **B. Keberfungsian Butir Diferensial (*Differential Item Functioning/DIF*)**

Dalam suatu alat ukur berupa tes, apabila terjadi keadaan dimana terdapat pengaruh dari faktor yang tidak diukur maka biasanya terjadi apa yang disebut dengan bias dari suatu tes. Istilah bias berarti tidak baik, bermakna rasial, menekan atau terlalu fanatik dengan objek yang akan diukur (Osterlind, 1983). Bias dalam tes merupakan suatu kondisi tes yang tidak adil (*unfair*), tidak konsisten, dan terkontaminasi oleh faktor-faktor di luar faktor yang hendak di tes. Jadi bias dalam tes mempunyai makna yang tidak mendukung sifat tes yang valid dan reliabel.

Dalam memaknai istilah bias dalam suatu tes terdapat dua macam pendekatan, sebagaimana dikemukakan oleh Osterlind (1983). Pertama, dengan melihat kegunaan skor

tes yang akan diperoleh. Kedua, pendekatan yang didasarkan pada pemahaman dari dugaan bahwa suatu tes itu bias. Pendekatan ini dilakukan dengan melihat butir-butir tes itu sendiri yakni adakah yang menyalahi kriteria-kriteria yang telah ditentukan. Pendekatan kedua ini dikenal dengan istilah bias internal butir dalam suatu tes.

Secara metodologis maksud analisis bias internal butir tes adalah untuk membedakan antara perbedaan sebenarnya dan bias dalam pengukuran. Perbedaan kelompok dalam performans tes atau pada butir-butir tes tidak dapat secara otomatis sebagai petunjuk adanya bias karena perbedaan skor mungkin merupakan refleksi dari perbedaan pengetahuan dan pengalaman kelompok. Oleh karena itu dikemukakan konsep kesulitan relatif dan bias dioperasionalkan sebagai kesulitan relatif butir berbeda yang berlebihan atau menyimpang untuk kelompok tertentu atau kelompok tetap. Hanya saja, jika suatu butir relatif lebih sulit untuk satu kelompok dan kesulitan tersebut tidak relevan terhadap konstruk tes maka butir tersebut dikatakan bias. Secara konseptual bias adalah interaksi antara performans butir dengan keanggotaan kelompok (Camilli & Shepard, 1994). Untuk menjaga perbedaan antara kesulitan relatif dan bias maka Holland dan Thayer dalam Camilli & Shepard (1994) serta Adams (1992) menyebut kesulitan relatif ini sebagai *Differential Item Functioning* atau disingkat *DIF*.

### **C. Metode Pendeteksian DIF**

*DIF* atau kesalahan sistematik butir-butir suatu tes dapat dideteksi lewat berbagai metode/prosedur. Satu asumsi penting guna penyelidikan bias adalah bahwa butir-butir tes tersebut mengukur hanya satu atribut ketrampilan pengambil tes atau *unidimensionality*. Tanpa asumsi ini interpretasi respons butir amat sangat kompleks. Asumsi itu benar jika pola respons butir dari sub kelompok yang diambil dari populasi yang sama diharapkan sama tanpa memperhatikan perbedaan kemampuan sub kelompok secara keseluruhan. Oleh karenanya seluruh strategi atau prosedur untuk mendeteksi bias berdasarkan asumsi ini.

Metode pendeteksian bias butir sangat berguna karena fokus terbesarnya adalah penelitian dengan cermat terhadap butir-butir tes secara individual dan untuk menguji kembali butir-butir tes tersebut apakah berfungsi seperti yang diharapkan. Dengan

demikian jelas bahwa prosedur pendeteksian bias internal hanya menjawab satu set pertanyaan tentang validitas untuk pengambil tes dari kelompok-kelompok yang berbeda.

Berbagai metode pendeteksian *DIF* pada suatu perangkat tes di antaranya sebagaimana dikemukakan Adams (1992) yaitu (1) analisis faktor, (2) pengujian daya pembeda butir dengan menggunakan *point biserial* dan korelasi parsial, (3) pengujian tingkat kesukaran butir dengan menggunakan berbagai transformasi, (4) metode anova, (5) metode *laten trait*, (6) pendekatan *chi square* yang menggunakan model *log linier*, dan (7) metode statistik Mantel-Haenszel. Sedangkan Osterlind (1983) mengemukakan lima teknik untuk mendeteksi *DIF* pada butir-butir tes yaitu (1) analisis variansi, (2) transformasi indeks kesukaran butir, (3) kai kuadrat, (4) analisis respons distraktor, dan (5) kurva karakteristik butir. Selanjutnya, Shealy & Stout (1993) mengembangkan prosedur baru untuk mendeteksi *DIF* yakni prosedur *SIBTEST* yang dikembangkan berdasarkan teori respons butir dan terbukti efektif digunakan dalam berbagai kondisi serta mampu mengatasi kesalahan tipe I. Prosedur ini terus berkembang dan bahkan dapat dimodifikasi tidak hanya mendeteksi *DIF* namun juga efek faktor penyebab timbulnya *DIF*. Penggunaan prosedur ini mensyaratkan subtes valid dan subtes yang diuji dalam pendeteksian *DIF*.

Metode lain yang juga berkembang saat ini adalah pendekatan fungsi respons harapan pada grafik *DIF* yang dikembangkan untuk melengkapi analisis *DIF* global tradisional (Lewis, 1985; Mislevy, Wingersky, dan Sheehan, 1994). Namun, metode ini belum banyak direkomendasikan oleh para ahli. Selanjutnya, Camilli & Shepard (1994) mengemukakan delapan teknik untuk mendeteksi *DIF*, namun lima di antaranya tidak direkomendasikan untuk digunakan. Kelima teknik tersebut adalah (1) transformasi indeks kesukaran butir; (2) penyesuaian transformasi indeks kesukaran butir; (3) prosedur aturan emas, (4) analisis variansi, dan (5) perbedaan pada korelasi biserial titik butir. Sedangkan tiga teknik lainnya yang direkomendasikan untuk digunakan adalah tabel kontingensi (metode statistik Mantel Haenszel), pendekatan regresi logistik, dan teknik kurva karakteristik berdasarkan teori respons butir.

#### **D. Pendeteksian *DIF* dengan Pendekatan Regresi Logistik.**

Pendekatan regresi logistik dikembangkan oleh Swaminathan & Rogers (1990) dengan menggunakan peserta tes sebagai unit analisis. Pendekatan ini mempunyai

persamaan umum matematis yang menggunakan koefisien kelompok, skor total, dan bentuk interaksi yang dihitung dan diuji signifikansinya dengan cara pembandingan model. Salah satu kelebihan metode ini, di antaranya dikemukakan Paul et al. (2004: 1) yang menyatakan bahwa pendekatan ini bersifat sangat fleksibel dalam penggunaannya. Secara umum, pendekatan regresi logistik untuk mendeteksi *DIF* dinyatakan dengan persamaan sebagai berikut (Swaminathan & Rogers, 1990):

$$P(u_i=1) = \frac{\exp \psi_i}{1 + \exp \psi_i} \quad (1)$$

dengan

$$\Psi_i = \delta + \tau_1 G_i + \tau_2 X_i + \tau_3 (G_i X_i) \quad (2)$$

Keterangan:

$i$  : Indeks peserta tes.

$u_i$  : Respons peserta tes  $i$  terhadap suatu butir (1= benar, 0= salah).

$X_i$  : Skor total peserta tes  $i$ .

$G_i$  : Kelompok peserta tes  $i$  berada (1= Focal, 2 = Reference).

$G_i X_i$  : Interaksi skor total peserta  $i$  dan kelompok peserta  $i$ .

$\tau_1$  : Gabungan log odds ratio.

$\tau_2$  : Perbedaan kemampuan antar peserta tes dalam memperoleh jawaban benar.

$\tau_3$  : Koefisien interaksi.

Koefisien-koefisien dalam model regresi logistik dihitung dengan menggunakan metode *likelihood maximum*. Hipotesis yang didasarkan pada pembandingan model berbeda diuji dengan statistik perbandingan likelihood. Pengujian hipotesis dengan menggunakan pendekatan regresi logistik untuk mendeteksi *DIF* dilakukan dengan menggunakan tiga model sebagai berikut:

a. Model I

Secara umum model I merupakan model yang paling lengkap dan dinyatakan dengan persamaan sebagai berikut :

$$\Psi_i = \delta + \tau_1 G_i + \tau_2 X_i + \tau_3 (G_i X_i) \quad (3)$$

b. Model II

Model II merupakan model yang lebih sederhana daripada model I. Artinya, model I mempunyai satu suku lebih banyak dari pada model II, dan model II dinyatakan dengan persamaan sebagai berikut :

$$\Psi_i = \delta + \tau_1 G_i + \tau_2 X_i \quad (4)$$

c. Model III

Model III merupakan model yang lebih sederhana lagi daripada model II dan dinyatakan dengan persamaan sebagai berikut:

$$\Psi_i = \delta + \tau_2 X_i \quad (5)$$

Adapun pengujian terhadap model-model tersebut dilakukan dengan menggunakan dua langkah sebagai berikut :

a. Langkah 1

Model I diuji dibandingkan dengan model II untuk menguji apakah efek interaksi  $G_i X_i$  diperlukan. Pembandingan ini dilakukan dengan membandingkan harga  $\chi^2_{hitung}$  dengan harga  $\chi^2_{tabel}$  pada derajat kebebasan 1 dan taraf signifikansi  $\alpha$ . Apabila efek interaksi ini terbukti tidak signifikan dilanjutkan langkah berikutnya.

b. Langkah 2

Pada langkah ini model II diuji dibandingkan dengan model III. Pengujian ini dilakukan untuk menguji apakah faktor  $G_i$  diperlukan untuk memprediksi respons butir dan dilakukan juga dengan derajat kebebasan 1 dan taraf signifikansi  $\alpha$ . Apabila pada langkah kedua ini terbukti signifikan maka mengindikasikan adanya *DIF* pada suatu butir tes yang dianalisis.

## E. Penutup

Demikian secara garis besar mengenai pendekatan regresi logistik dalam pendeteksian *DIF*. Langkah-langkah pendeteksian *DIF* menggunakan pendekatan regresi logistik ini dapat dilakukan dengan menggunakan program *SPSS for windows*. Lebih lanjut sesuai dengan kemajuan IPTEKS, berbagai temuan yang muncul dari analisis empirisnya akan semakin memperbaiki penggunaan pendekatan ini. Semoga bermanfaat.

## DAFTAR PUSTAKA

- Adams, R.J. (1992). Item Bias. Dalam J.P. Keeves, (Ed). *The IEA technical handbook*. The Hague: The International Association for the Evaluation of Educational Achievement (IEA).
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publication.
- Mislevy, R.J., & Wingersky, M. S., & Sheehan, K. M. (1994). *Dealing with uncertainty about item parameters: Expected response functions* (Research Report RR-94-28-ONR). Princeton, NJ: Educational Testing service.
- Osterlind, S. J. (1983). *Test item bias*. Beverly Hills, CA: Sage Publication.
- Paul, C., Gerald, V.B., & Eric, L. (2004). *Differential item functioning in a cognitive test*. Diambil pada tanggal 27 April 20034 dari <https://www.alz.washington.edu/NONMEMBER/KENTUCKY0502/CRANE/tsld011>
- Shealy, R., & Stout, W.F. (1993). An item response theory model for test bias. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 197-239). Hillsdale, NJ: Erlbaum.
- Swaminathan , H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.