

Penerapan Regresi Logistik pada Jawaban Siswa terhadap Butir Tes dengan Penskoran Dikotomi

Oleh :
Heri Retnawati
Pendidikan Matematika FMIPA UNY
(Riwati@yahoo.com)

ABSTRAK

Dalam dunia pendidikan, sebagian besar tes yang digunakan adalah tes pilihan ganda dengan cara penskoran dikotomi. Pada data yang dikotomi ini dapat dibangun model regresi logistik, yang melibatkan parameter kemampuan dan parameter butir tes. Model ini dapat digunakan untuk mengestimasi parameter kemampuan siswa maupun parameter butir tes. Ada tiga model regresi logistic yang dapat digunakan yakni model satu, dua dan tiga parameter.

Pendahuluan

Untuk mengetahui keberhasilan pelaksanaan pendidikan, biasanya dilakukan tes. Menurut Cronbach (Heri Retnawati, 2003: 19-20), tes merupakan suatu prosedur sistematis untuk mengamati dan menggambarkan satu atau lebih karakteristik seseorang dengan suatu skala numerik atau sistem kategorik. Berdasarkan hal ini, tes memberikan informasi yang bersifat kualitatif dan kuantitatif.

Tes dapat diklasifikasikan dengan beberapa macam, tergantung dari tujuannya (Anastasi dan Urbina, 1997 : 2-4). Tes prestasi belajar merupakan suatu bentuk tes untuk mendapatkan data, yang merupakan informasi untuk melihat seberapa banyak pengetahuan yang telah dimiliki dan dikuasai oleh seseorang sebagai akibat dari pendidikan dan pelatihan (Anastasi dan Urbina, 1997: 42-43). Berdasarkan informasi yang diperoleh ini, pada proses seleksi, siswa dapat dikelompokkan sesuai dengan kemampuannya, yang diterima atau tidak diterima. Hal ini sesuai dengan fungsi tes prestasi seperti yang dikemukakan Gronlund (1976: 16), yang menyatakan bahwa tes prestasi berfungsi sebagai alat untuk penempatan, fungsi formatif, fungsi diagnostik dan fungsi sumatif.

Berdasarkan bentuknya, tes prestasi belajar dapat dikelompokkan menjadi tiga jenis, yaitu : 1) objektif, yang sederhana terdiri dari bentuk jawaban singkat, benar-salah atau dua pilihan, dan menjodohkan, serta objektif pilihan ganda dengan alternatif jawaban lebih dari

dua, 2) uraian (Gronlund, 1976: 144). Pada proses penyekorannya, tes objektif merupakan data yang bersifat dikotomi. Dikotomi ini berarti bahwa ketika siswa menjawab butir dengan benar, diberikan skor 1 sedangkan jika salah diberikan skor 0. Dengan pendekatan teori tes klasik, dari skor dikotomi ini, kemampuan siswa dinyatakan dengan skor total menjawab benar. Cara ini bisa jadi kurang adil, karena tingkat kesulitan yang dimiliki tiap butir berbeda-beda.

Pada tulisan ini akan dikaji pendekatan lain dalam mengetahui kemampuan siswa dalam penskoran dikotomi, yakni dengan menerapkan regresi logistik.

Pembahasan

Dalam evaluasi yang dilaksanakan dalam pendidikan, siswa menjawab butir soal suatu tes yang berbentuk pilihan ganda dengan benar, biasanya diberi skor 1 dan 0 jika menjawab salah. Pada penyekoran dengan pendekatan teori tes klasik, kemampuan siswa dinyatakan dengan skor total yang diperolehnya. Prosedur ini kurang memperhatikan interaksi antara setiap orang siswa dengan butir.

Pendekatan teori respons butir merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Ada dua prinsip yang digunakan pada pendekatan ini, yakni prinsip relativitas dan prinsip probabilitas. Pada prinsip relativitas, unit dasar dari pengukuran bukanlah siswa atau butir, tetapi lebih kepada performance siswa relative terhadap butir. Jika β_n merupakan indeks dari kemampuan siswa ke n pada trait yang diukur, dan δ_i merupakan indeks dari tingkat kesulitan dari butir ke-i relative yang terkait dengan kemampuan yang diukur, maka bukan β_n atau δ_i yang merupakan unit pengukuran, tetapi lebih kepada perbedaan antara kemampuan dan dari siswa relative terhadap tingkat kesulitan butir atau $(\beta_n - \delta_i)$ perlu dipertimbangkan. Sebagai alternativenya perbandingan antara kemampuan terhadap tingkat kesulitan dapat digunakan. Jika kemampuan dari siswa melampaui tingkat kesulitan butir, maka respons siswa diharapkan benar, dan jika kemampuan siswa kurang dari tingkat kesulitan butir, maka respons siswa diharapkan salah (Keeves dan Alagumalai, 1999:24).

Pada teori respons butir, prinsip probabilitas menjadi perhatian. Misalkan kemampuan siswa ke n dinyatakan dengan θ_n dan tingkat kesulitan dari butir dinyatakan

dengan Δ_i maka sesuai dengan prinsip relativitas, jika $\theta_n > \Delta_i$ siswa diharapkan menjawab dengan benar, dan $\theta_n < \Delta_i$ siswa diharapkan menjawab salah. Lebih jauh lagi, jika kemungkinan (*odds*) dari respons siswa terhadap butir diberikan oleh $\frac{\theta_n}{\Delta_i} > 1$ (siswa diharapkan menjawab dengan benar), $\frac{\theta_n}{\Delta_i} < 1$, siswa diharapkan menjawab salah, dan $\frac{\theta_n}{\Delta_i} = 1$, akan terjadi jika kesempatan 50% menjawab benar.

Jika p_{ni} merupakan peluang menjawab benar, maka $1 - p_{ni}$ merupakan respons tidak benar dan odds untuk respons diberikan oleh

Probabilitas respons menjawab benar berada pada rentang 0 sampai dengan 1.0 dan hal ini menghalangi data dinyatakan sebagai skala interval. Skor mentah yang dihasilkan dari cara ini sulit dinyatakan sebagai skala. Untuk mengatasi permasalahan ini, dapat digunakan transformasi logistik, yang melibatkan logaritma natural dari odds.

yang senilai dengan

Misalkan $\ln \theta_n = \beta_n$ dan $\Delta_i = \delta_i$ maka

Persamaan tersebut senilai dengan

atau dengan kata lain probabilitas respons menjawab benar ($x_{ni} = 1$) dapat dituliskan sebagai

dan probabilitas respons menjawab salah

Model ini merupakan model logistik univariat (Hosmer dan Lemeshow, 1989).

Bentuk persamaan yang lebih dikenal dalam pengukuran untuk model ini, yang biasa disebut dengan model Rasch (Hambleton, Swaminathan, dan Rogers, 1991: 12) sebagai berikut :

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

secara acak dapat menjawab butir i dengan benar

θ : tingkat kemampuan subyek (sebagai variabel bebas)

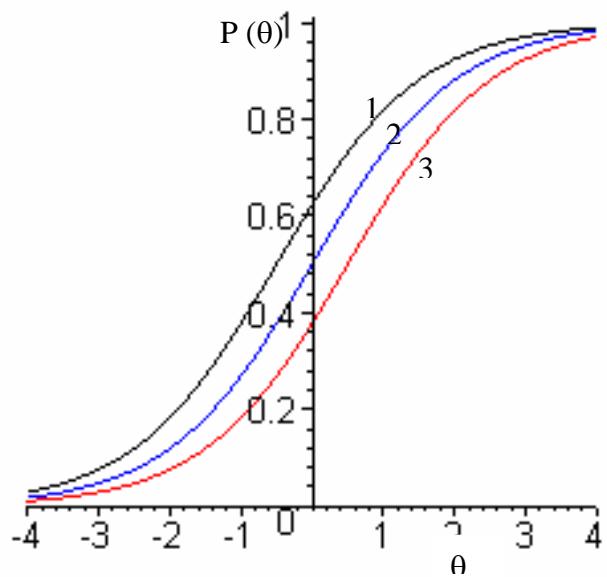
b_i : indeks kesukaran butir ke- i

e : bilangan natural yang nilainya mendekati 2,718

n : banyaknya butir dalam tes

Parameter b_i merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Misalkan suatu butir tes mempunyai parameter $b_i = 0,3$, artinya diperlukan kemampuan minimal 0,3 pada skala untuk dapat menjawab benar dengan peluang 50%. Semakin besar nilai parameter b_i , maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b_i , maka makin sulit butir soal tersebut.

Hubungan peluang menjawab benar $P_i(\theta)$ dengan tingkat kemampuan peserta (θ) dapat digambarkan sebagai kurva karakteristik butir (*item characteristic curve, ICC*). Gambar 1 berikut merupakan ilustrasi kurva karakteristik butir untuk model Rasch (1 parameter, 1P), dengan butir 1 ($b=-0,5$), butir 2 ($b=0$) dan butir 3($b=0,5$).

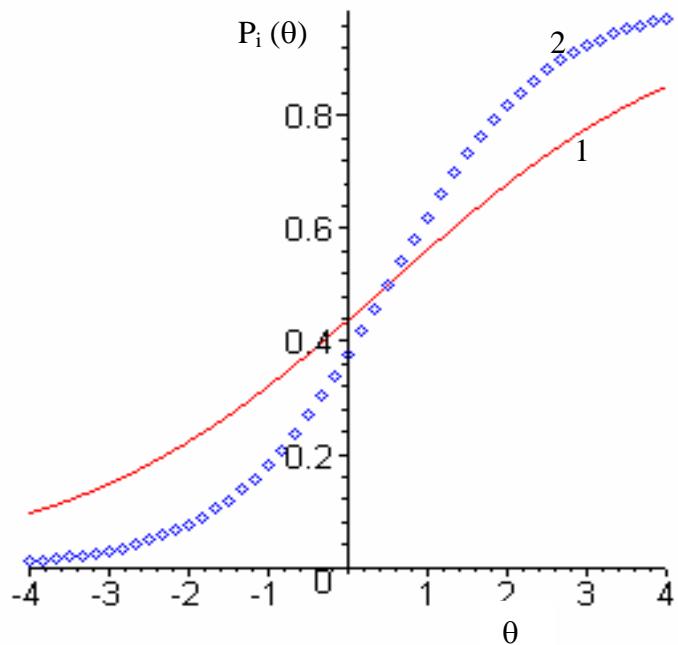


Gambar 1. Kurva karakteristik butir untuk model 1P, dengan butir 1 ($b=-0,5$), butir 2 ($b=0$) dan butir 3($b=0,5$).

Jika $(\theta - b_i)$ ditransformasi menjadi $a_i(\theta - b_i)$ dengan a_i suatu konstanta, maka a_i ini merupakan tingkat daya pembeda butir (*item difficulty*). Selanjutnya model ini merupakan model logistik 2 parameter (2P) dengan parameter butir yaitu indeks kesukaran butir (b_i) dan indeks daya beda butir (a_i), yang memenuhi :

Jika a_i bernilai 1, maka model 2 parameter ini menjadi model logistik 1 parameter.

Sebagai ilustrasi, kurva karakteristik butir 1 ($a=0,5$; $b=0,5$) dan butir 2 ($a=1$; $b=0,5$).

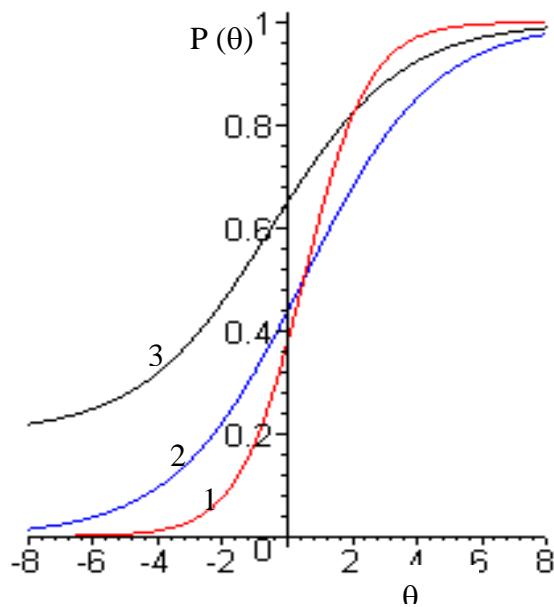


Gambar 2. Kurva karakteristik butir model 2P, dengan butir 1 ($a=0,5$; $b=0,5$) dan butir 2 ($a=1$; $b=0,5$)

Pada gambar 2 tersebut, asimtot kiri (untuk $\theta \rightarrow -\infty$) adalah 0. Jika bukan 0, maka nilai ini merupakan parameter tebakan semu (*pseudo guessing*) atau c_i , sehingga model logistik menjadi model 3 parameter (3P). Dengan adanya tebakan semu pada model logistik tiga parameter, memungkinkan siswa yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Secara matematis, model logistik tiga parameter dapat dinyatakan sebagai berikut (Hambleton, Swaminathan, dan Rogers, 1991: 17, Hambleton, dan Swaminathan, 1985 : 49, Van der Linden dan Hambleton, 1997: 13).

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta-b_i)}}{1+e^{a_i(\theta-b_i)}} \dots \dots \dots \quad (10)$$

Sebagai ilustrasi, gambar 3 merupakan kurva karakteristik butir 1 ($a=1$, $b=0,5$, $c=0$), butir 2($a=0,5$, $b=0,5$, $c=0$) dan butir 3 ($a=0,5$, $b=0,5$, $c=0,2$).



Gambar 3. kurva karakteristik butir model 3P, dengan butir 1 ($a=1, b=0,5, c=0$), butir 2($a=0,5, b=0,5, c=0$) dan butir 3 ($a=0,5, b=0,5, c=0,2$)

Nilai kemampuan peserta (θ) terletak di antara -4 dan $+4$, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Secara teoretis, nilai b_i terletak di antara $-\sim$ dan $+\sim$. Suatu butir dikatakan baik jika nilai ini berkisar antara -2 dan $+2$ (Hambleton dan Swaminathan, 1985: 107). Jika nilai b_i mendekati -2 , maka indeks kesukaran butir sangat rendah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesukaran butir sangat tinggi untuk suatu kelompok peserta tes.

Parameter a_i merupakan daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik, a_i merupakan kemiringan (*slope*) dari kurva di titik b_i pada skala kemampuan tertentu. Karena merupakan kemiringan, diperoleh semakin besar kemiringannya, maka semakin besar daya pembeda butir tersebut. Secara teoretis, nilai a_i ini terletak antara $-\sim$ dan $+\sim$. Pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dengan kemampuan yang diukur, dan a_i terletak antara 0 dan 2 (Hambleton dan Swaminathan, 1985: 37).

Peluang menjawab benar dengan memberikan jawaban tebakan semu dilambangkan dengan c_i , yang disebut dengan *tebakan semu*. Parameter ini memberikan suatu kemungkinan asimtot bawah yang tidak nol (*nonzero lower asymptote*) pada kurva karakteristik butir (ICC). Parameter ini menggambarkan probabilitas peserta dengan

kemampuan rendah menjawab dengan benar pada suatu butir yang mempunyai indeks kesukaran yang tidak sesuai dengan kemampuan peserta tersebut. Besarnya harga c_i diasumsikan lebih kecil daripada nilai yang akan dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir. Pada suatu butir tes, nilai c_i ini berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai c_i tidak lebih dari $1/k$, dengan k banyaknya pilihan (Hullin, 1983: 36).

Fungsi informasi butir (*Item Information Functions*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes, dan pembandingan beberapa perangkat tes. Fungsi informasi butir menyatakan kekuatan atau sumbangannya butir tes dalam mengungkap latent trait yang diukur dengan tes tersebut. Dengan fungsi informasi butir diketahui butir yang mana yang cocok dengan model sehingga membantu dalam seleksi butir tes. Secara matematis, fungsi informasi butir memenuhi persamaan sebagai berikut.

keterangan :

i : 1,2,3,...,n

$I_i(\theta)$: fungsi informasi butir ke- i

$P_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

$P'_i(\theta)$: turunan fungsi $P_i(\theta)$ terhadap θ

$Q_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton dan Swaminathan, 1985: 94). Berhubungan dengan hal ini, fungsi informasi perangkat tes akan tinggi jika butir tes mempunyai fungsi informasi yang tinggi pula. Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

Nilai-nilai indeks parameter butir dan kemampuan peserta merupakan hasil estimasi. Karena merupakan hasil estimasi, maka kebenarannya bersifat probabilitas dan tidak terlepas dari kesalahan pengukuran. Dalam teori respon butir, kesalahan penaksiran

standar (*Standard Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil atau sebaliknya (Hambleton, Swaminathan dan Rogers, 1991, 94). Jika nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ dan nilai estimasi *SEM* dinyatakan dengan $\hat{SEM}(\hat{\theta})$, maka hubungan keduanya, menurut Hambleton, Swaminathan, dan Rogers (1991 : 94) dinyatakan dengan

$$\text{SEM}(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad \dots \dots \dots \quad (13)$$

Pada model 1P, 2P dan 3P, untuk menganalisis jawaban siswa, yang perlu menjadi perhatian adalah pengestimasian parameter butir dan parameter kemampuan peserta. Dalam pengestimasian ini, dikenal fungsi likelihood. Fungsi likelihood untuk kasus dengan N siswa dan n butir dapat dinyatakan dengan

$$L(\theta, b; u) = \prod_i \prod_j P_i(\theta_j; b_i)^{u_{ij}} [P_i(\theta_j; b_i)]^{1-u_{ij}} \quad (1P)$$

$$L(\theta, a, b, c; u) = \prod_i \prod_j P_i(\theta_j; a_i, b_i, c_i)^{u_{ij}} \left[P_i(\theta_j; a_i, b_i, c_i) \right]^{1-u_{ij}} \quad (3P)$$

Selanjutnya diestimasi nilai-nilai yang memaksimumkan fungsi ini. Prosedur yang dapat dipilih yakni prosedur likelihood maksimum gabungan (*joint maximum likelihood, JML*) atau prosedur likelihood maksimum marginal (*marginal maximum likelihood, MML*) atau juga dengan pendekatan Bayes.

Untuk mengestimasi parameter-parameter butir pada model logistik 1P, 2P maupun 3P, ada beberapa perangkat lunak yang dapat digunakan, diantaranya Rascal (1P), Ascal (2P dan 3P), Bilog, Xalibrate dan Multilog. Keluaran (*output*) dari program-program ini juga menyediakan hasil pengestimasi parameter peserta tes.

Langkah selanjutnya adalah mengetahui kecocokan model dari data yang dianalisis. Uji statistik untuk kecocokan model salah satunya uji perbandingan likelihood (*likelihood ratio test*). Uji ini digunakan untuk mengecek apakah estimasi parameter butir dalam grup skor yang berbeda bernilai sama pada kesalahan penyampelan dari estimasi. Secara teoritis, responden yang berukuran N dapat dibuat menjadi interval-interval pada skala kontinum

untuk θ , yang merupakan dasar untuk mengestimasi nilai θ . Statistik Khi-kuadrat dari perbandingan *likelihood* digunakan untuk membandingkan frekuensi menjawab benar dan tidak benar dari respons pada interval yang diharapkan dari model yang cocok pada rata-rata interval $\hat{\theta}_h$, dengan persamaan :

dengan n_g merupakan banyaknya interval, r_{hj} merupakan frekuensi respons yang benar untuk butir pada interval h , N_h merupakan banyaknya anggota sampel yang berada dalam interval, dan $P_j(\bar{\theta}_h)$ merupakan nilai dari fungsi respons sesuai model untuk butir j pada $\bar{\theta}_h$, yang merupakan kemampuan rata-rata responden pada interval h (Mislevy dan Bock, 1990).

Pada program Bilog, untuk menentukan banyaknya interval yang dibuat pada skala kontinu untuk θ , mula-mula dibuat maksimum 20 interval. Setiap responden disarangkan pada interval tersebut termasuk estimasi EAP (*expected a posteriori*), berdasarkan tipe prior yang dispesifikasikan pemakai dari skor yang diperoleh responden. Pada setiap butir tes, probablilitas harapan dari respons yang sesuai dengan estimasi rata-rata EAP untuk kemampuan dari kasus yang berada dalam interval digunakan sebagai proporsi harapan untuk interval tersebut.

Khi-kuadrat perbandingan kemungkinan dihitung setelah mengkombinasikan interval-interval yang ekstrim, hingga frekuensi harapan pada gabungan interval-interval tersebut lebih dari lima. Derajat kebebasan dari khi-kuadrat perbandingan kemungkinan sama dengan banyaknya interval-interval yang telah dikombinasikan (Mislevy dan Bock, 1990).

Sebagai gambaran penerapan analisis regresi logistik pada data dikotomi sebagai jawaban dari peserta tes terhadap suatu kumpulan butir sebagai berikut. Sebuah data dikotomi yang dibangkitkan dari kemampuan siswa yang berdistribusi $N(0,1)$, yang terdiri dari 300 peserta tes dan 20 butir soal. Selanjutnya data dianalisis dengan model logistik 3P dengan bantuan program Bilog. Hasil analisis sebagai berikut.

Tabel 1. Hasil analisis data dikotomi hasil simulasi dengan model 3P

Butir	a	b	c	χ^2	dk	p-value
1	0.612	1.451	0.271	3.7	4	0.44
2	1.97	1.57	0.054	6.3	4	0.17
3	1.458	-1.646	0.248	0.2	1	0.63
4	1.411	-0.933	0.285	0.4	2	0.81
5	2.05	0.254	0.091	4.2	2	0.11
6	0.598	-0.965	0.295	4.3	4	0.36
7	2.252	-0.75	0.202	1.1	1	0.29
8	0.891	1.91	0.302	4.6	4	0.33
9	0.368	2.278	0.5	4.4	4	0.35
10	0.644	-0.421	0.253	1.6	3	0.65
11	1.102	0.488	0.13	3.2	3	0.36
12	1.005	1.725	0.138	3.6	4	0.45
13	0.517	1.286	0.384	1.6	5	0.9
14	1.351	-1.275	0.258	5.3	1	0.02
15	1.071	1.962	0.185	2.4	4	0.66
16	1.546	1.438	0.174	5.4	4	0.24
17	1.176	0.192	0.132	2.7	3	0.43
18	1.777	-0.588	0.212	7.2	2	0.02
19	0.835	2.032	0.237	2.4	4	0.66
20	0.98	0.977	0.111	7.3	4	0.11

Berdasarkan tabel tersebut, dengan taraf signifikansi 1%, keseluruhan butir cocok dengan model 3P. Adapun nilai fungsi informasi untuk tes sebesar 5,701 pada level kemampuan 1.6208. Dengan rumus (13) dapat diperoleh kesalahan pengukuran sebesar 2,388.

Kesimpulan

Pada penskoran tes dengan skor dikotomi, dapat digunakan regresi logistik, yang menggunakan prinsip relativitas dan probabilitas. Model ini dapat digunakan untuk mengestimasi parameter kemampuan siswa maupun parameter butir tes. Ada tiga model regresi logistic yang dapat digunakan yakni model satu parameter (tingkat kesulitan butir), dua parameter (tingkat kesulitan dan daya pembeda butir) dan tiga parameter (tingkat kesulitan, daya pembeda dan tebakan semu butir).

Referensi

Anastasi, A. & Urbina,S. (1997). *Psychological testing*. Upper Saddle River, NJ : Prentice Hall.

Gronlund, N.E. (1976). *Measurement and evaluation in teaching*. New York : Macmillan Publishing Co.

Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.

Heri Retnawati. (2003). Keberfungsiang butir diferensial pada perangkat tes seleksi masuk SMP. *Tesis*. Universitas Negeri Yogyakarta, tidak dipublikasikan.

Hosmer, D.W. dan Lemeshow,S. (1989). *Applied Logistic Regressions*. New York : John Willwy and Sons.

Hullin, C. L., et al. (1983). *Item response theory : Application to psychological measurement*. Homewood, IL : Dow Jones-Irwin.

Keeves, J.P. dan Alagumalai,S. (1999). New approaches to measurement. Dalam Masters, G.N. dan Keeves, J.P.(Eds). *Advances in measurement in educational research and assesment*. Amsterdam : Pergamon.

Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. New York : Hold, Rinehart and Wiston,Inc.

Mislevy, R.J. & Bock,R.D. (1990). *BILOG 3 : Item analysis & test scoring with binary logistic models*. Moorseville : Scientific Sofware Inc.

Van der Linden, W.J. dan Hambleton, R.K. (1997). Item response theory:brief history, common models and extention. Dalam Van der Linden, W.J. dan Hambleton, R.K. (Eds). *Handbook of item response theory*. New York : Springer.