

ANALISIS KARAKTERISTIK PARAMETER BUTIR BERDASARKAN MODEL LOGISTIK 3 PARAMETER

Dina Huriaty¹

1. STKIP PGRI Banjarmasin
dina.huriaty@stkipbjm.ac.id

ABSTRAK

Tujuan penelitian ini adalah mengidentifikasi karakteristik parameter butir tes berdasarkan model logistik tiga parameter. Penelitian ini adalah penelitian deskriptif kuantitatif dengan menggunakan data respons ujian nasional tingkat SMP yang berisi 40 butir soal bentuk pilihan ganda dengan empat pilihan jawaban. Data respons Ujian Nasional merupakan data dokumentasi dengan jumlah sampel respons 2500 respons siswa yang dipilih secara acak. Jumlah ini memenuhi syarat ukuran sampel minimal untuk model logistik tiga parameter. Pemilihan model logistik tiga parameter didasari bahwa pada tes-tes kemampuan yang menggunakan format pilihan ganda adalah contoh situasi dimana model logistik tiga parameter cocok untuk digunakan. *Software* yang digunakan untuk mengestimasi dan analisis hasil penelitian ini adalah *BILOG-MG*. Hasil penelitian menunjukkan statistik parameter butir memperlihatkan nilai rata-rata indeks daya beda butir tergolong baik, rata-rata indeks kesukaran butir sedang, dan rata-rata parameter *pseudo guessing* baik. Estimasi terhadap indeks reliabilitas menunjukkan instrument tes reliable.

Kata Kunci: karakteristik, parameter butir, model logistik tiga parameter

PENDAHULUAN

Item Response Theory atau teori respons butir merupakan teori tentang bagaimana variabel responden dan variabel butir menentukan data respons ketika seseorang menjawab butir tersebut (Umar, 1999). Menurut teori respons butir, performansi peserta tes diasumsikan sebagai fungsi dari suatu kemampuan laten. Model ini memprediksikan probabilitas individu pada kemampuan atau nilai *latent-trait* tertentu. Terdapat tiga unsur dalam teori respons butir, yaitu butir tes, peserta tes, dan respons peserta tes. Ketiga unsur ini saling berhubungan dan menghasilkan fungsi respons atau kurva karakteristik butir. Pada hubungan ini, karakteristik peserta tes dinyatakan dalam bentuk θ , sedangkan karakteristik butir dinyatakan dengan parameter butir a (daya beda), b (indeks kesukaran), dan c (tebakan semu), dimana respon peserta tes terhadap butir dinyatakan dalam bentuk probabilitas jawaban benar $P(\theta)$. Hubungan ini dituliskan sebagai $P_j(\theta) = \varphi(\theta, a_j, b_j, c_j)$. Kalibrasi butir adalah proses estimasi terhadap parameter-parameter tersebut dan menempatkannya pada skala umum.

Pada tes-tes kemampuan yang menggunakan format pilihan ganda seperti format soal ujian adalah contoh situasi dimana model logistik tiga parameter cocok untuk digunakan. Peserta tes cenderung untuk memilih jawaban terbaik yang mereka anggap paling menarik jika mereka tidak dapat menemukan jawabannya, sehingga faktor menebak dipertimbangkan dalam penelitian ini.

Pertanyaan yang muncul adalah bagaimana karakteristik butir-butir tes yang dianalisis dengan menggunakan model logistik tiga parameter? Tujuan penelitian ini adalah mengidentifikasi karakteristik parameter butir tes berdasarkan model logistik tiga parameter. Hasil penelitian ini diharapkan dapat memberikan sumbangan teori bagi perkembangan di bidang pengukuran dan pengujian, terutama analisis butir menggunakan teori respons butir dengan model logistik tiga parameter.

Teori respons butir sebagai teori tes modern menjadi kerangka kerja statistik untuk menyelesaikan masalah pengukuran, seperti pengembangan tes (Hambleton & Jones, 1993). Penyusunan tes dengan pendekatan IRT menempatkan parameter soal bersifat invarian, artinya fungsi atau kurva dari respons butir tidak berubah, walaupun kelompok peserta yang menjawab butir yang sama itu berubah-ubah. Tingkat kesulitan butir dan kemampuan peserta tes diukur dengan skala yang sama. Teori respons butir mempunyai kelebihan dibandingkan Teori tes klasik, yaitu statistik butir tidak tergantung pada kelompok, skor tes yang diperoleh dapat menggambarkan kemampuan individu, tidak memerlukan tes yang paralel untuk menghitung koefisien reliabilitas, dan dapat menyediakan ukuran yang tepat untuk setiap skor kemampuan.

Manfaat utama pendekatan IRT adalah (1) parameter butir bersifat invarian, maksudnya fungsi/kurva respons butir tidak berubah; dan (2) seleksi butir-butir berdasarkan atas banyaknya informasi butir dan informasi tes (Hambleton, Swaminathan, & Rogers, 1991:91). Manfaat potensial penggunaan pendekatan IRT dalam konstruksi tes, antara lain (1) proses kalibrasi butir dengan *sample free*, dan (2) pencapaian pengukuran orang dengan *item free* (Green, Yen, & Burket, 1989).

Teori respons butir didasarkan pada dua postulat, yaitu: (1) kemampuan setiap individu pada suatu butir soal dapat diperkirakan dengan seperangkat faktor yang disebut karakteristik laten (*latent-traits*), (2) hubungan antara kemampuan individu pada suatu butir dan perangkat kemampuan yang mendasarinya sesuai dengan grafik fungsi monoton, yang disebut kurva karakteristik butir (*ICC= Item Characteristic Curve*). Kurva karakteristik butir menunjukkan bahwa semakin tinggi tingkat kemampuan individu, semakin besar peluang menjawab benar suatu butir.

Ada tiga asumsi dalam teori respons butir, yaitu unidimensi, independensi lokal, dan invariansi parameter. Selain memenuhi persyaratan unidimensi, independensi lokal, dan invariansi parameter, hal lain dalam teori respons butir yang juga perlu diperhatikan adalah penentuan model respons atau karakteristik butir. Teori respons butir menggunakan pendekatan probabalistik untuk menyatakan hubungan antara kemampuan individu dengan harapan menjawab benar. Model distribusi yang digunakan adalah distribusi logistik.

Ada tiga model logistik dalam teori respons butir, yaitu model logistik satu, dua, dan tiga parameter. Model-model ini sesuai untuk data respons butir yang diskor dikotomis (Hambleton, Swaminathan, & Rogers, 1991: 12). Yang membedakan ketiga model ini adalah banyaknya parameter yang digunakan untuk menggambarkan karakteristik butir pada setiap model logistiknya atau parameter-parameter butir. Parameter-parameter butir tersebut adalah indeks kesukaran butir, indeks daya beda butir, dan tebakan semu. Jika ciri individu dinyatakan sebagai

parameter θ , sedangkan ciri butir dinyatakan sebagai parameter butir a , b , dan c . Respons individu terhadap butir dinyatakan dalam bentuk probabilitas jawaban yang benar $P(\theta)$; maka hubungan ini untuk butir ke- j , dapat dinyatakan sebagai berikut:

$$P_j(\theta) = \varphi(\theta, a_j, b_j, c_j)$$

$P_j(\theta)$ menyatakan probabilitas jawaban benar untuk butir ke- j , parameter θ menyatakan ciri individu yang dapat berupa kemampuan akademik peserta tes, parameter a adalah ciri butir yang menunjuk pada daya beda butir, parameter b adalah ciri butir yang berkenaan dengan kesukaran butir, dan parameter c adalah ciri butir yang menyatakan faktor tebakan semu (*pseudo-guessing*). Untuk mengestimasi parameter-parameter pada teori respons butir diperlukan suatu ukuran banyaknya peserta yang cukup besar. Ukuran banyaknya peserta untuk mengestimasi parameter tergantung dari model yang dipakai. Model tiga parameter memerlukan peserta tes yang lebih besar dibandingkan jika menggunakan model dua parameter atau satu parameter.

Pada model logistik tiga parameter, probabilitas peserta tes dalam menjawab benar suatu butir ditentukan oleh semua parameter butir, yaitu parameter kesukaran butir, parameter daya beda butir, dan *pseudo-guessing* (tebakan semu). Model logistik tiga parameter dinyatakan pada persamaan berikut.

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Keterangan:

$P_i(\theta)$: probabilitas peserta tes dengan kemampuan θ yang menjawab benar butir i .

θ : tingkat kemampuan peserta tes

b_i : indeks kesukaran butir ke- i

a_i : indeks daya beda butir ke- i

c_i : indeks tebakan semu butir ke- i

e : bilangan natural = 2,718

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

D : faktor penskalaan = 1,7

n : banyaknya butir tes.

(Hambleton, et al. ,1991; Embretson & Reise, 2000; Partchev, 2004).

Pada model logistik tiga parameter, asimtot bawah pada kurva karakteristik butir mungkin saja tidak sama dengan nol. Hal ini menunjukkan bahwa individu yang mempunyai kemampuan rendah masih mempunyai peluang untuk menjawab benar suatu butir.

Model logistik tiga parameter adalah model yang paling umum dari ketiga model. Dengan kurva berbentuk S dan asimptut yang lebih rendah, model ini sangat tepat ketika individu dengan kemampuan rendah terkadang dapat merespons dengan benar butir yang sulit (Hulin, Drasgow, & Parsons, 1983). Pada penelitian ini data dianalisis dengan menggunakan model logistik tiga parameter. Tes-tes kemampuan yang menggunakan format pilihan ganda dan instrumen sikap adalah contoh situasi dimana model logistik tiga parameter cocok

untuk digunakan. Peserta tes cenderung untuk memilih jawaban terbaik yang mereka anggap paling menarik jika mereka tidak dapat menemukan jawabannya. Sehingga faktor menebak dipertimbangkan dalam penelitian ini.

Batas-batas untuk parameter-parameter butir pada model logistik tiga parameter dalam mengestimasi parameter butir adalah 0,30 hingga 2,0 untuk parameter a (daya beda), untuk parameter b (kesukaran butir) dapat merupakan bilangan riil, tetapi secara normal estimasi b berada pada rentang -3,0 dan 3,0. Asimptot dibatasi pada interval $0 < c < 1$. Estimasi c (tebakan semu) seringkali mendekati $1/m$ untuk butir-butir pada tes kemampuan yang berbentuk pilihan ganda, dengan m adalah banyaknya pilihan jawaban (Hulin, Drasgow & Parsons, 1983: 35-36). Pada pengukuran sikap, nilai-nilai c yang bukan nol jarang terjadi, kecuali untuk kecenderungan respons yang sifatnya umum.

Salah satu langkah penting dalam estimasi butir pada teori respons butir adalah pemilihan model. Beberapa pertimbangan dalam memutuskan untuk memilih model teori respons butir yang tepat untuk digunakan adalah: (1) ketersediaan sampel peserta tes dalam jumlah yang cukup, (2) sifat dari sampel, dan (3) ketersediaan program komputer yang digunakan untuk mengestimasi kemampuan dan parameter-parameter butir (Hari Setiadi & Hambleton, 1996:14). Menurut Hambleton & Swaminathan (1985), pertimbangan dalam memilih model respons butir yang digunakan tergantung pada karakteristik data. Beberapa pertimbangan dalam memilih model adalah: (1) model logistik satu parameter digunakan jika daya pembeda dari populasi sama; (2) model logistik satu dan dua parameter dipilih jika indeks *guessing* minimal; (3) model logistik satu, dua, dan tiga parameter dipilih jika tes yang diadministrasikan bukan tes kecepatan; dan (4) memenuhi asumsi unidimensi

METODE PENELITIAN

Penelitian ini adalah penelitian deskriptif kuantitatif dengan menggunakan data respons ujian nasional mata pelajaran Matematika tingkat SMP yang berisi 40 butir soal bentuk pilihan ganda dengan empat pilihan jawaban. Data respons Ujian Nasional merupakan data dokumentasi dengan respons dipilih secara acak. Jumlah respons yang menjadi sampel adalah 2500 respons siswa. Jumlah ini telah memenuhi syarat ukuran sampel minimal jika menggunakan model logistik tiga parameter. Model logistik yang dipilih adalah model logistik tiga parameter. Hal ini didasari bahwa pada tes-tes kemampuan yang menggunakan format pilihan ganda seperti format soal ujian nasional adalah contoh situasi dimana model logistik tiga parameter cocok untuk digunakan. Peserta tes cenderung untuk memilih jawaban terbaik yang mereka anggap paling menarik jika mereka tidak dapat menemukan jawabannya, sehingga faktor menebak menjadi dipertimbangkan.

Software yang digunakan sebagai alat bantu dalam estimasi dan analisis hasil penelitian ini adalah *BILOG-MG*. Menurut Crocker & Algina (1986: 354), Hambleton (1991: 42-50), dan Yen & Fitzpatrick (2006: 131-132), program komputer untuk estimasi kebolehjadian maksimum adalah program *Bilog*. *Bilog-MG* fit untuk model logistik satu, dua, dan tiga parameter Program *Bilog-MG*

Analisis Karakteristik Parameter Butir
Berdasarkan Model Logistik 3 Parameter

dapat mengestimasi butir-butir bentuk pilihan ganda, dan dapat digunakan untuk skala besar dalam mengestimasi kemampuan latent.

HASIL DAN PEMBAHASAN

Hasil estimasi terhadap parameter butir digunakan untuk mengetahui karakteristik butir dan peserta tes. Hasil analisis dilakukan terhadap output Bilog phase 1, phase 2, dan phase 3. Analisis butir pada output Bilog phase 1 meliputi tingkat kesukaran butir dan indeks korelasi biserial. Analisis terhadap tingkat kesukaran butir pada output Bilog phase 1, terdapat 16 butir yang mudah, 24 butir yang tingkat kesukarannya sedang, dan tidak ada butir yang tergolong sukar . Hal ini dapat dilihat pada Tabel 1.

Tabel 1. Tingkat Kesukaran Butir

Indeks Tingkat Kesukaran	Kriteria	Nomor Butir
$0,00 \leq p < 0,30$	Sukar	---
$0,30 \leq p \leq 0,70$	Sedang	2, 3, 4, 5, 9, 10, 12, 13, 14, 15, 16, 17, 18, 22, 25, 26, 28, 29, 31, 34, 35, 36, 37, 38
$0,70 < p \leq 1,00$	Mudah	1, 6, 7, 8, 11, 19, 20, 21, 23, 24, 27, 30, 32, 33, 39, 40

Hasil analisis terhadap indeks korelasi biserial diperoleh 2 butir soal memiliki indeks korelasi biserial kurang dari 0,30, yaitu butir soal nomor 3 dengan indeks korelasi sebesar 0,233 dan butir soal nomor 28 dengan indeks korelasi sebesar 0,211. Selengkapnya dapat dilihat pada Tabel 2.

Tabel 2. Indeks Korelasi Biserial Butir

Nomor Butir	Indeks Korelasi Biserial	Nomor Butir	Indeks Korelasi Biserial
1	0,663	21	0,699
2	0,618	22	0,587
3	0,233	23	0,688
4	0,631	24	0,679
5	0,556	25	0,542
6	0,439	26	0,327
7	0,482	27	0,615

8	0,656	28	0,211
9	0,569	29	0,669
10	0,548	30	0,543
11	0,742	31	0,460
12	0,663	32	0,483
13	0,602	33	0,324
14	0,692	34	0,621
15	0,532	35	0,554
16	0,584	36	0,473
17	0,429	37	0,357
18	0,688	38	0,388
19	0,609	39	0,649
20	0,683	40	0,573

Analisis kecocokan butir terhadap model logistik tiga parameter juga dapat dilihat pada output Bilog phase 2 ini. Kecocokan butir terhadap model dilihat berdasarkan nilai distribusi χ^2 dan derajat kebebasan (*db*) masing-masing butir. Suatu butir cocok dengan model jika jika $\chi_{hit}^2 \leq \chi_{tab}^2$ atau $p > 0,01$.

Proses estimasi parameter butir kemudian dilakukan setelah mengeliminasi butir-butir yang memiliki indeks korelasi biserial kurang dari 0,30. Estimasi terhadap 38 butir dilakukan setelah mengeliminasi butir nomor 3 dan 28 menghasilkan indeks korelasi biserial lebih dari 0,30 pada semua butir, sehingga analisis dapat diteruskan ke output Bilog phase 2. Dengan menetapkan $\alpha = 1\%$, dan membandingkan dengan output Bilog phase 2, Hasil analisis menunjukkan terdapat 89,47% butir soal yang fit atau cocok dengan model logistik tiga parameter dan 4 butir tidak cocok dengan model, yaitu butir nomor 9, 23, 25, dan 38.

Hasil analisis kecocokan model menunjukkan bahwa tes fit dengan model logistik tiga parameter, karena terdapat lebih dari 80% butir yang cocok dengan model. Tes mempunyai parameter daya beda butir yang tergolong baik, yaitu berada pada interval [0,30, 2,0]. Butir yang daya beda yang sangat tinggi terdapat pada butir soal nomor 17 yaitu 1,599. Indeks daya beda terkecil terdapat pada butir soal nomor 33, yaitu 0,481; demikian juga halnya dengan indeks kesukaran butir. Indeks kesukaran butir berada pada interval [-3, 3], butir soal yang mempunyai indeks kesukaran terkecil adalah butir soal nomor 33, yaitu -2,274. Indeks kesukaran butir paling besar terdapat pada butir soal nomor 38, yaitu 1,189. Terdapat 12 butir memiliki indeks parameter *pseudoguessing* yang kurang baik karena nilai *c* lebih dari 0,25.

Analisis terhadap output Bilog phase 3 memberikan deskripsi tentang nilai rata-rata hitung, simpangan baku, estimasi indeks reliabilitas, dan standar baku pengukuran (SEM), dapat dilihat pada Tabel 3.

Tabel 3. Statistik Estimasi Skor

Statistik	Nilai
Mean	-0.0034
Simpangan Baku	0,9711
Indeks Reliabilitas	0,9041
SEM	1,2546

Tabel 3 memperlihatkan bahwa estimasi indeks reliabilitas adalah 0,9041. Nilai estimasi ini lebih dari 0,70, hal ini berarti bahwa instrumen reliable.

PENUTUP

Berdasarkan hasil penelitian dan pembahasan, diperoleh kesimpulan, yaitu statistik parameter butir memperlihatkan nilai rata-rata indeks daya beda butir tergolong baik, rata-rata indeks kesukaran butir sedang, dan rata-rata parameter *pseudo guessing* baik. Berdasarkan estimasi terhadap indeks reliabilitas menunjukkan instrumen tes reliable.

Implikasi hasil penelitian ini berkenaan dengan pengembangan soal. Setiap pelaksanaan penilaian di sekolah sebaiknya terlebih dahulu melakukan uji coba terhadap instrumen tes yang akan diujikan. Hal ini dilakukan untuk dapat mengetahui kualitas instrumen tes yang dikembangkan. Hasil penelitian ini diharapkan dapat menjadi rujukan sebagai salah satu alternative model dalam melakukan analisis butir soal. Informasi tentang kualitas butir soal dapat memudahkan untuk perencanaan dan pelaksanaan tes berikutnya.

DAFTAR PUSTAKA

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston Inc.
- du Toit, M. (Ed.) (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publisher.
- Green, D.R., Yen, W.M., & Burket, G.R. (1989). Experiences in the application of item response theory in test construction. *Journal of Educational Measurement*, 2, 297-312.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and Applications*. Boston, MA: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Hambleton, R. K & Jones, R. W. (1993). *An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development*. Diambil pada tanggal 5 Juli 2010, dari www.ncme.org/pubs/items/10.pdf.
- Hari Setiadi & Hambleton, R. K. (1996, June). *Item banks to improve assessment practices*. Internal Paper. University of Massachusetts Amherst.

- Hulin, C.L., Drasgow, F. & Parsons, C.K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.
- Partchev, I. (2004). *A visual guide to item response theory*. Diambil pada tanggal 12 April 2009.
- Umar, J. (1999). Item banking. Dalam G.N. Masters & J.P. Keeves (Eds.). *Advances in measurement in educational research and assessment* (pp. 207-218) Oxford: Elsevier Science Ltd.
- van der Linden, W. J. & Hambleton, R. K. (Eds). (1997). *Handbook of modern item response theory*. New York: Springer.
- Yen, W. M & Fitzpatrick, A. R. (2006). Item response theory dalam R.L. Brennan (Ed.), *Educational measurement. 4th ed.* (pp. 111-153). Westport, CT: American Council on Education and Praeger Publishers.