

**ANALISIS BUTIR SOAL
DENGAN TEORI TES KLASIK (*CLASSICAL TEST THEORY*) DAN
TEORI RESPONS BUTIR (*ITEM RESPONSE THEORY*)
(Studi Kasus: Soal Ujian Olimpiade Sains Provinsi Bidang Informatika 2009)**

Ida Mariati Hutabarat
Jurusan Matematika FMIPA Universitas Cenderawasih

Abstrak

Olimpiade Sains Provinsi (OSP) sangat penting untuk menumbuhkan suasana kompetitif di kalangan pelajar agar bersaing secara sehat dalam penguasaan ilmu pengetahuan dan teknologi. Perangkat tes OSP harus valid, reliabel, dan terbebas dari bias. Penelitian ini bertujuan untuk membandingkan model IRT (*Item Response Theory*) dengan indikator pada teori uji klasik (*Classical Test Theory*) dan memilih model IRT yang sesuai untuk menggambarkan karakteristik butir soal dan kemampuan peserta ujian. Data yang digunakan adalah data dari Olimpiade Sains Provinsi bidang Informatika Tahun 2009, sebanyak 60 item pilihan ganda dengan mengambil sampel sebanyak 500 peserta. Tahapan yang dilakukan adalah pendugaan karakteristik butir soal dan kemampuan peserta, serta pemeriksaan kesesuaian model. Proses analisis dilakukan dengan menggunakan *software* TAP dan Bilog-MG. Berdasarkan hasil analisis, terdapat perbedaan parameter daya pembeda dan tingkat kesukaran pada metode teori uji klasik dan teori respons butir. Daya pembeda dan tingkat kesukaran pada teori uji klasik dipengaruhi oleh kemampuan kelompok sedangkan pada teori respons butir dipengaruhi oleh kemampuan individu. Berdasarkan teori uji klasik, tingkat kesukaran dari soal OSP yang diujikan tergolong sedang, dengan reliabilitas skor tes sebesar 0,910 yang menunjukkan tingkat ketepatan dan kekonsistenan peserta dalam menjawab soal sudah sangat baik. Model yang paling sesuai untuk menggambarkan butir-butir soal pada soal OSP adalah model teori respons butir tiga parameter logistik (IRT 3PL), yaitu 85% butir soal sudah sesuai atau dapat digambarkan oleh model.

Kata kunci : Teori Respons Butir, Teori Uji Klasik, Olimpiade Sains Provinsi (OSP)

PENDAHULUAN

Olimpiade Sains Provinsi (OSP) sangat penting untuk menumbuhkan suasana kompetitif di kalangan pelajar agar bersaing secara sehat dalam penguasaan ilmu pengetahuan dan teknologi. Pelaksanaan olimpiade secara berkelanjutan akan berdampak positif pada pelaksanaan proses pembelajaran sehingga menjadi lebih kreatif dan inovatif.

Pada gilirannya, siswa akan memiliki kesempatan mengembangkan seluruh aspek kepribadian dan kemampuannya melalui pembelajaran yang aktif, kreatif, efektif dan menyenangkan (PAKEM).

OSP merupakan langkah awal bagi siswa-siswi yang berbakat yang nantinya diharapkan terus mengikuti olimpiade di jenjang yang lebih tinggi. Secara umum OSP bertujuan meningkatkan mutu pendidikan secara komprehensif melalui penumbuhkembangan budaya belajar, kreativitas dan motivasi meraih prestasi terbaik melalui kompetisi yang sehat serta menjunjung nilai-nilai sportivitas. Melalui olimpiade ini diharapkan akan terjadi upaya peningkatan kompetensi yang tentu saja dibarengi dengan upaya peningkatan mutu pembelajaran di sekolah.

Pengukuran kemampuan anak diperlukan perangkat tes yang berkualitas sehingga kemampuan kognitif siswa dapat diungkapkan. Kualitas sebuah perangkat tes dapat dilihat dengan melakukan analisis kualitatif dan kuantitatif. Analisis kualitatif merupakan analisis yang dilakukan sebelum tes diberikan kepada peserta tes dengan melihat kesesuaiannya dengan aspek materi, konstruksi dan bahasa, sedangkan analisis kuantitatif dapat dilakukan dengan menggunakan dua pendekatan yaitu teknik teori uji klasik (*classical test theory*) dan teori respons butir (*item response theory*).

Analisis butir soal dengan teori uji klasik merupakan yang termudah meskipun memiliki beberapa keterbatasan. Beberapa aspek yang diperhatikan dalam teori uji klasik yaitu tingkat kesukaran butir, daya pembeda butir, penyebaran pilihan jawaban, dan reliabilitas skor tes (Safari, 2000). Teori respons butir merupakan teori pengukuran modern yang biasanya digunakan dalam analisis butir soal. Dalam teori ini digunakan model matematis untuk menghubungkan karakteristik butir soal dengan kemampuan responden. Hubungan tersebut digambarkan melalui kurva karakteristik butir.

Oleh karena itu, penulis menelaah apakah item-item soal OSP bidang Informatika tahun 2009 tersebut benar-benar dapat mengukur apa yang seharusnya diukur memiliki tingkat kesulitan yang kurang lebih sepadan dengan kemampuan peserta tes, indeks daya beda yang tinggi, serta faktor tebakan (*guessing*) yang seminimal mungkin sehingga dapat memberikan informasi pengukuran yang akurat.

Tujuan Penelitian

Tujuan dalam penelitian ini adalah :

1. Membandingkan model IRT (*Item Response Theory*) dengan indikator pada teori uji klasik (*Classical Test Theory*)
2. Mencari model yang sesuai untuk menggambarkan butir-butir soal pada soal OSP bidang Informatika.

Manfaat Penelitian

Bila tujuan penelitian di atas tercapai, manfaat yang diperoleh antara lain:

1. Sebagai masukan dalam penyusunan soal olimpiade, khususnya olimpiade bidang Informatika.
2. Sebagai sumbangan teoretik dalam ranah pengujian alat ukur, khususnya model IRT dan CTT.

DATA DAN METODE

Data

Data yang digunakan adalah hasil jawaban OSP bidang Informatika. Ukuran sampel yang diambil sebanyak 500 siswa. Jumlah butir soal sebanyak 60 butir soal pilihan ganda, dimana setiap soal memiliki lima pilihan jawaban yaitu A, B, C, D, dan E. Soal terdiri dari dua bagian utama, yaitu nomor 1-30 adalah soal analitik, dan nomor 31-60 adalah soal algoritmika, yaitu pengetahuan pemrograman berbasis bahasa Pascal. Jika dirinci lebih lanjut berdasarkan kisi-kisi OSP bidang Informatika, soal terdiri dari tiga materi uji, yaitu aritmatika, logika, dan aspek teknis (Tabel 1).

Tabel 1. Materi butir soal

No.	Aspek Kognitif	Item Soal
1.	Aritmatika	1,2,3,4,5,6,7,8,9,14,16,17,18,23
2.	Logika	10,11,12,13,15,19,20,21,22,24,25,26,27, 28,29,30,35,36,41,42,44,45
3.	Teknis praktis	37,38,39,40,43,46,47,48,49,50,51,52,53, 54,55,56,57,58,59,60

Metode

Tahapan-tahapan yang dilakukan pada penelitian ini adalah :

1. Teori Uji Klasik

a) Menghitung indeks kesukaran

Rumus untuk menentukan besarnya indeks kesukaran secara matematis dirumuskan oleh Saifuddin (2003) sebagai berikut:

$$P = \frac{n_1}{N}$$

dengan P adalah indeks kesukaran butir, n_1 = jumlah peserta tes yang menjawab benar, dan N = banyaknya siswa yang menjawab butir soal tersebut.

Menurut Allen & Yen (1979) tingkat kesukaran dikategorikan sebagai berikut :

Tabel 2. Kategori butir soal

Nilai p	Kategori
$p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang
$p > 0,7$	Mudah

b) Menghitung daya pembeda

Teknik untuk menentukan nilai daya beda dengan menggunakan teknik korelasi phi (ϕ). sebagai berikut:

$$\phi = \frac{P_H - P_L}{2\sqrt{(p)(q)}}$$

dengan

ϕ = angka indeks diskriminasi phi yang dianggap sebagai angka indeks diskriminasi butir

p_H = proporsi orang yang menjawab benar kelompok atas

p_L = proporsi orang yang menjawab benar kelompok bawah

p = proporsi seluruh peserta tes yang menjawab benar dan

$q = 1 - p$

Menurut Ebel & Frisbie (1986) memberikan patokan indeks daya beda sebagai berikut:

Tabel 3. Indeks daya beda

Indeks daya	Evaluasi butir
0,4 ke atas	Butir yang sangat baik
0,3 – 0,39	Sedikit atau tidak memerlukan revisi
0,2 – 0,29	Butir memerlukan revisi
< 0,19	Butir harus dieliminasi

c) Menghitung reliabilitas soal

Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan kekonsistenan skor tes. Indeks reliabilitas berkisar antara 0-1. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), makin tinggi pula ketepatannya.

Reliabilitas dapat dihitung dengan koefisien alfa, dalam Crocker dan Algina (1986) didefinisikan sebagai berikut :

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_x^2} \right)$$

dengan

n = jumlah butir soal

σ_x^2 = ragam skor total

σ_i^2 = ragam skor per butir soal

Reliabilitas tes soal bentuk pilihan ganda dapat juga digunakan rumus Kuder Richadson 20 (KR-20) sebagai berikut ini :

$$KR - 20 = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^n p(1-p)}{\sigma_x^2} \right)$$

dengan

k = jumlah butir soal

σ_x^2 = ragam skor total

p = proporsi jawaban benar terhadap semua jawaban peserta tes

d) Penyebaran Pilihan Jawaban

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan soal. Hal ini dimaksudkan untuk mengetahui berfungsi tidaknya jawaban yang tersedia. Safari (2000) menyatakan bahwa suatu pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila paling tidak dipilih oleh 5% peserta ujian dan lebih banyak dipilih oleh kelompok siswa yang belum paham materi.

2. Teori Respons Butir (IRT)

a) Menghitung parameter kemampuan setiap model.

Ada tiga model logistik yang sering digunakan saat ini (Hambleton *et al.*, 1991) yaitu

1. Model logistik satu parameter (model rasch) atau *item response theory 1-parameter logistic* (IRT 1PL) yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran. Kurva karakteristik butir soal untuk model satu parameter diberikan oleh persamaan :

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

2. Model logistik dua parameter (model rasch) atau *item response theory 2-parameter logistic* (IRT 2PL) yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal. Kurva karakteristik butir soal untuk model dua parameter diberikan oleh persamaan :

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

3. Model logistik tiga parameter (model rasch) atau *item response theory 3-parameter logistic* (IRT 3PL) yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran, daya pembeda soal dan peluang menebak (*guessing*). Kurva karakteristik butir soal untuk model tiga parameter diberikan oleh persamaan :

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

dengan

$P_i(\theta)$ = peluang bahwa peserta tes dengan kemampuan menjawab butir soal ke-i dengan benar

a_i = parameter daya pembeda soal butir ke-i

b_i = parameter tingkat kesukaran, yaitu satu titik pada skala *ability* di mana kemungkinan untuk menjawab benar sebesar 0,5

c_i = peluang tebakan benar butir ke-i

Q = parameter kemampuan peserta tes

D = faktor penskalaan yang diikutkan untuk menjadikan fungsi logistik serupa mungkin dengan fungsi ogive normal ($D = 1,702$)

b) Menghitung parameter karakteristik butir soal untuk setiap model.

c) Mencari model yang paling sesuai untuk menggambarkan setiap soal

Proses analisis dilakukan dengan menggunakan software TAP dan Bilog-MG. Program TAP digunakan untuk analisis teori tes klasik dan Program Bilog-MG digunakan untuk menganalisis model IRT 1PL, IRT 2PL, dan IRT 3PL.

HASIL DAN PEMBAHASAN

Nilai statistik skor siswa dalam Tes OSP bidang Informatika tahun 2009 dapat dilihat pada Tabel 4 di bawah ini :

Tabel 4. Kategori butir soal

Statistik	Skor Nilai Siswa
Rataan skor	23,26
Standar Deviasi	10,648
Skor Minimum	2
Skor Maksimum	55
Median	23

Teori Uji Klasik

Berdasarkan klasifikasi tingkat kesukaran soal menurut Allen & Yen (1979), persentase dan penyebaran soal tergolong sukar, sedang dan mudah dapat dilihat pada Tabel 5 di bawah ini

Tabel 5. Tingkat kesukaran soal

Kategori	Nomor Soal	Total	Persentase
Sukar	9,10,12,15,19,22,27,29,30,31,35,36, 45,47,48,49,55,58,59,60	20	33
Sedang	1,2,3,4,5,6,7,8,11,17,18,21,23,25,26,28,32,34,37, 38,39,40,41,42,43,44,46,50,51,52,53,54,56,57	34	57
Mudah	13,14,16,20,24,33	6	10
Total		60	100

Dari tabel di atas diketahui persentase soal tergolong sukar 33%, sedang 57 % dan susah 10%. Jadi soal tes yang diujikan masuk dalam kategori sedang,

Daya pembeda soal dapat dihitung dengan nilai korelasi poin biserial atau daya pembeda dengan mengambil data 25% dari atas dan 25% dari bawah. Persentase soal yang tergolong baik, cukup baik, perlu direvisi, dan soal yang belum bisa membedakan (harus dieliminasi) masing-masing sebesar 47%, 18%, 22%, dan 13%. Penyebaran evaluasi butir dapat dilihat pada Tabel 6 di bawah ini.

Tabel 6. Daya Pembeda Soal

Evaluasi butir	Nomor Soal	Total	Persentase
Butir yang sangat baik	4,6,18,22,32,33,34,37,38,39,40,41, 42,43,44,45,46,47,49,50,51,52,53, 54,55,56,57,58	28	47
Sedikit atau tidak memerlukan revisi	1,3,5,8,9,10,11,16,17,31,59	11	18
Butir memerlukan revisi	2,7,13,14,19,20,23,24,26,28,36,48,60	13	22
Butir harus dieliminasi	12,15,21,25,27,29,30,35	8	13
Total		60	100

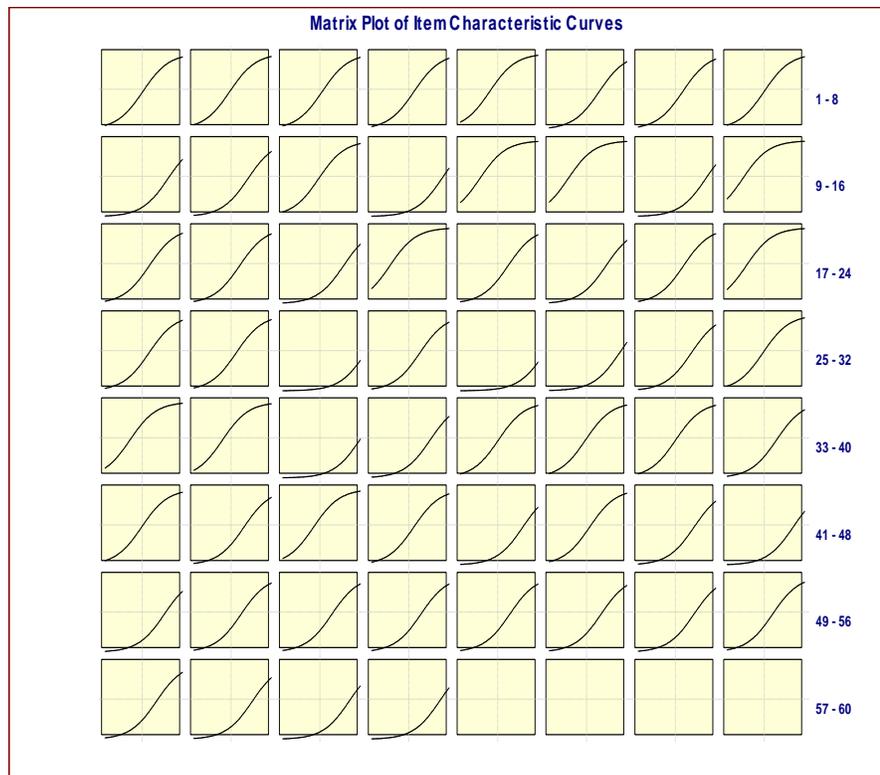
Berdasarkan analisis dari sebaran pilihan jawaban, setiap pilihan jawaban dipilih oleh peserta ujian sehingga dapat disimpulkan bahwa pilihan jawaban berfungsi sebagai pengecoh. Namun ada beberapa soal yang proporsi setiap pilihan jawaban hampir merata yaitu soal nomor : 6, 9, 10, 16, 19, 22, 30, 31, 35, 45, 48, 49, 59, 60.

Reliabilitas skor tes dilihat menggunakan koefisien alfa. Nilai koefisien alfa yang diperoleh dari hasil program TAP sebesar 0,910 yang menunjukkan tingkat ketepatan dan kekonsistenan peserta dalam menjawab soal sudah sangat baik.

Model Teori Respons Butir

Satu parameter Logistik (IRT 1PL)

Hasil pendugaan parameter butir setiap soal ujian OSP menggunakan model IRT 1PL mempunyai nilai b (tingkat kesukaran) yang cukup beragam, yaitu berkisar antara -1,794 sampai 3,504. Pada model 1PL nilai a untuk setiap soalnya dianggap sama yaitu 0,58. Melalui nilai estimasi tingkat kesukaran butir ada 15% butir soal yang tidak baik yaitu nomor 12, 15, 27, 29, 30, 35, 48, 59, 60, karena dianggap terlalu sukar oleh peserta tes.

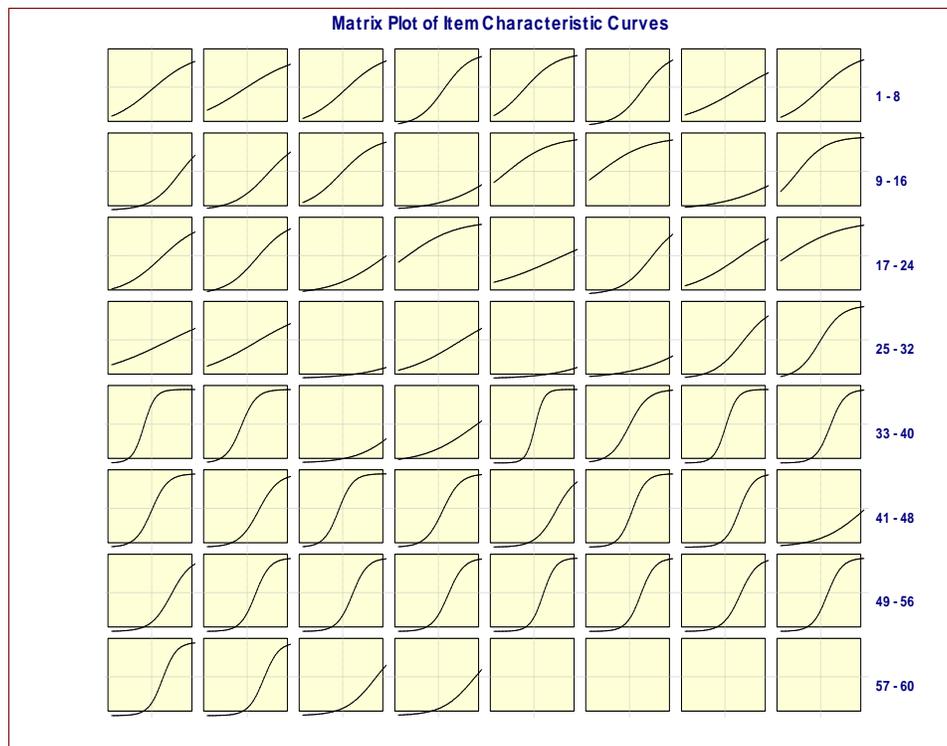


Gambar 1. Kurva karakteristik butir soal 1- 60 OSP untuk Model IRT 1PL

Berdasarkan uji kesesuaian model dan kurva karakteristik butir pada Gambar 1. menunjukkan sebesar 58,3% butir soal belum sesuai atau belum dapat digambarkan oleh model. Soal-soal yang belum dapat digambarkan adalah soal nomor 1, 2, 7, 12, 13, 15, 21, 23, 24, 25, 26, 27, 28, 29, 30, 33, 34, 36, 37, 38, 39, 40, 41, 43, 44, 46, 47, 50, 51, 52, 53, 54, 56, 57, 58, 60. Berdasarkan hasil analisis di atas, lebih dari 50% butir soal yang belum sesuai (tidak dapat digambarkan oleh model) sehingga model IRT 1PL belum dapat menggambarkan butir-butir soal OSP.

Satu parameter Logistik (IRT 2PL)

Hasil pendugaan parameter butir soal untuk model IRT 2PL diperoleh nilai a (daya beda) berkisar dari nilai 0,203 sampai dengan 1,881. Nilai b (tingkat kesukaran) berkisar dari nilai -2,669 sampai dengan 6,88.

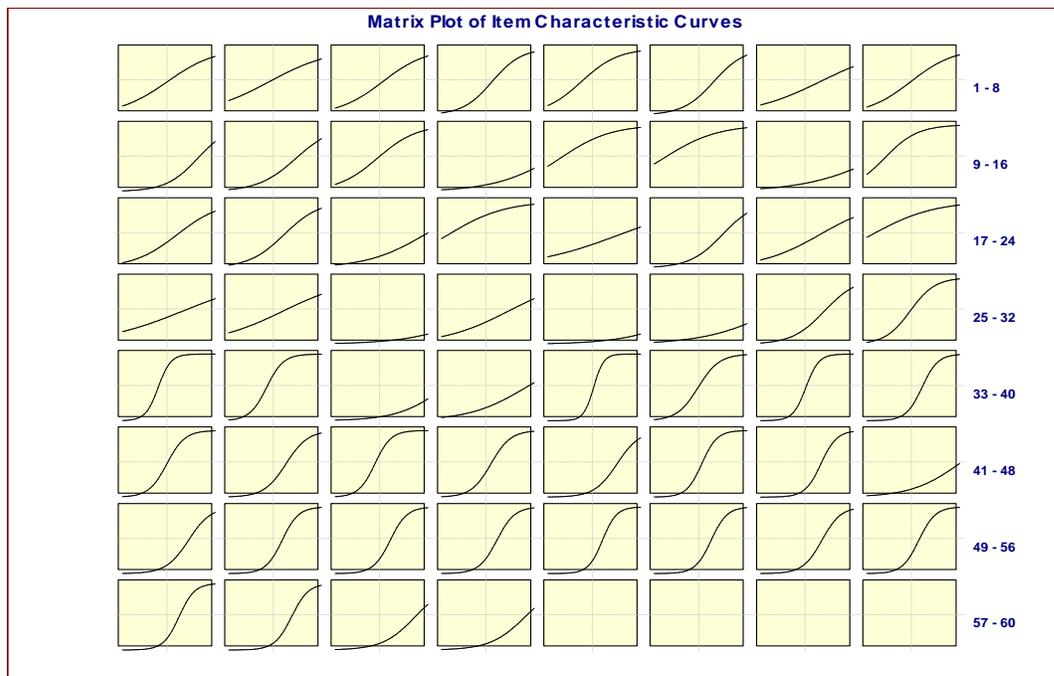


Gambar 2. Kurva karakteristik butir soal 1- 60 OSP untuk Model IRT 2PL

Berdasarkan uji kesesuaian model dan kurva karakteristik butir pada Gambar 2. menunjukkan sebesar 23,3% butir soal belum sesuai atau belum dapat digambarkan oleh model. Soal-soal yang belum dapat digambarkan adalah soal nomor 12, 13, 14, 15, 19, 20, 23, 24, 27, 29, 30, 35, 36, 48. Berdasarkan hasil analisis di atas, persentase soal yang sesuai lebih banyak daripada soal yang tidak sesuai. Hal ini menunjukkan parameter daya pembeda berpengaruh dalam model, sehingga disimpulkan model IRT 2PL sudah cukup baik untuk menggambarkan butir-butir soal OSP.

Tiga parameter Logistik (IRT 3PL)

Hasil pendugaan parameter butir soal untuk model IRT 3PL diperoleh nilai a (daya beda) berkisar dari nilai 0,209 sampai dengan 0,441 . Nilai b (tingkat kesukaran) berkisar dari nilai -2,52 sampai dengan 3,604, dan nilai c (peluang menebak) berkisar dari nilai 0,00 sampai dengan 0,106. Pada karakteristik peluang tebakan benar diperoleh 100% memiliki estimasi parameter yang baik dengan nilai estimasi kurang dari 0,2. Hal ini berarti butir soal tidak memberikan peluang kepada peserta yang berkemampuan rendah untuk menjawab benar dengan cara menebak.



Gambar 3. Kurva karakteristik butir soal 1- 60 OSP untuk Model IRT 3PL

Berdasarkan uji kesesuaian model dan kurva karakteristik butir pada Gambar 3. menunjukkan sebesar 85% butir soal sesuai atau dapat digambarkan oleh model dan 15% belum dapat digambarkan dengan model. Soal-soal yang belum dapat digambarkan adalah soal nomor 13, 14, 19, 20, 23, 24, 29, 36, 48. Berdasarkan hasil analisis di atas, persentase soal yang sesuai lebih banyak daripada soal yang tidak sesuai. Hal ini menunjukkan parameter daya pembeda dan peluang menebak berpengaruh dalam model, sehingga disimpulkan model IRT 3PL sudah cukup baik untuk menggambarkan butir-butir soal OSP.

KESIMPULAN

Berdasarkan data penelitian serta pembahasan yang telah diuraikan tadi maka penulis menarik kesimpulan:

1. Terdapat perbedaan parameter daya pembeda dan tingkat kesukaran pada metode teori uji klasik dan teori respons butir. Daya pembeda dan tingkat kesukaran pada teori uji klasik dipengaruhi oleh kemampuan kelompok sedangkan pada teori respons butir dipengaruhi oleh kemampuan individu.
2. Model yang paling sesuai untuk menggambarkan butir-butir soal pada soal OSP adalah model teori respons butir tiga parameter logistik (IRT 3PL), yaitu 85% butir soal sudah sesuai atau dapat digambarkan oleh model. Butir soal yang baik adalah soal nomor.

DAFTAR PUSTAKA

- Aiken, L.R. 1987. *Assesment of Intelectual functioning*. Massachussetts: Allyn and Bacon
- Allen, M.J., & Yen, W.M. 1979. *Introduction to measurement theory*. California: Brookd/Cole Publishing Company.
- Crocker, L. & Algina, J 1986. *Introduction to Classical and Modern Tes Theory*. New York: Rinehart and Winston.
- Ebel, R.L., & Frisbie, D.A. 1986. *Essentials of educational measurement*. New Jersey: Prentice Hall.

Analisis Butir Soal dengan Teori Tes Klasik ... (Ida Mariati Hutabarat)

Hambleton, R. K., Swaminathan, H. & H. J. Rogers. 1991. *Fundamentals of Item Response Theory*. California: Sage Publications.

Nitko, A.J. 1996. *Educational Assesment of Students, Second Edition*. Ohio: Prentice Hall.

Saifuddin Azwar. 2003. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.

Safari. 2000. Panduan Analisis Butir Soal. <http://www.dikmenum.go.id/> [12 Oktober 2009]