

## Daya Beda, Tingkat Kesulitan, dan Tebaan Tes Biologi Kelas 8 Semester Gasal

### Discrimination, Difficulty, and Guessing The Biology Test 8<sup>th</sup> Grade By The Period Of The Odd Term

Suwarto

Universitas Veteran Bangun Nusantara,  
Jl. Sujono Humardani No. 1 Jombor Sukoharjo Jawa-Tengah, Sukoharjo, Indonesia  
Corresponding Email: suwartowarto@yahoo.com

**Abstract:** The aims of this research: (1) Want to know different effort in biology test items at 8 class odd semester, (2) want to know difficult level in biology test items at 8 class odd semester, (3) want to know guessing biology test items at 8 class odd semester. Biology test 8 class odd semester includes 100 question items and have tried on 200 respondent. For analysis is used R programming. The results show: (1) Different effort in biology test items at 8 class odd semester about 1,002 up to 2,969. The lowest different effort item is 8 item and the highest different effort item is 38 item. (2) Difficult level in biology test items at 8 class odd semester about 1,005 un to 1,992. The lowest difficult level test item is 83 and 93 item and the highest difficult level is 45 item. (3) Guessing biology test items at 8 class odd semester. Biology test 8 class odd semester about 0,018 up to 0,989. The lowest guessing is 80 item and the highest guessing is 68 item.

**Keywords:** Discrimination, Difficulty, and Guessing

## 1. PENDAHULUAN

Setiap semester sekolah lanjutan tingkat pertama (SLTP) selalu mengadakan ujian semester. Ujian semester ini menggunakan tes sebagai alat ukurnya. Tes yang baik akan memberikan hasil ukur yang baik. Menurut Gronlund & Linn (1990:5) *Test an instrument or systematic procedure for measuring a sample of behavior. (Answers the question "How well does the individual perform---either in comparison with others or in comparison with a domain of performance tasks?")*. Cronbach (1970:26) mendefinisikan sebuah tes: *a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category-system*. Rusli (1988:4) tes adalah seperangkat pertanyaan yang dibuat untuk diberikan kepada siswa dengan syarat-syarat tertentu, atau tes adalah prosedur yang sistematis untuk mengobservasi tingkah laku. Gronlund & Lin, Cronbach, dan Rusli sejalan dalam mendefinisikan tes. Mereka mendefinisikan tes yang dipandang dari sudut prosedur pengukuran. Prosedur pengukuran dilakukan secara sistematis untuk mendapatkan informasi tingkah laku siswa.

Teori Respons Butir. 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 kemampuan siswa relatif terhadap butir. Jika  $\beta_n$  merupakan indeks dari kemampuan siswa ke n pada trait yang diukur, dan  $\delta_i$  merupakan indeks dari tingkat kesulitan dari butir ke-i relative yang terkait dengan kemampuan yang diukur, maka bukan  $\beta_n$  atau  $\delta_i$  yang merupakan unit pengukuran, tetapi lebih kepada perbedaan antara kemampuan dan dari siswa relatif terhadap tingkat kesulitan butir atau ( $\beta_n - \delta_i$ ) perlu dipertimbangkan. Sebagai alternatifnya 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  $\beta_n$  dan tingkat kesulitan dari butir dinyatakan dengan  $\delta_i$  maka sesuai dengan prinsip relativitas, jika  $\beta_n > \delta_i$  siswa diharapkan menjawab dengan benar, dan  $\beta_n < \delta_i$  siswa diharapkan menjawab salah. 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, digunakan transformasi logistik, sehingga hubungan antara tingkat kesulitan butir dan peluang menjawab benar bukan hubungan linear. Dalam teori respons butir, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir. Ini berarti bahwa peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta yang mempunyai kemampuan rendah.

Hambleton & Swaminathan (1985:16) dan Hambleton, Swaminathan, & Rogers (1991:9) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu unidimensi, independensi lokal dan invariansi parameter. Ketiga asumsi dapat dijelaskan sebagai berikut. Unidimensi, artinya setiap butir tes hanya mengukur satu kemampuan. Contohnya, pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan siswa dalam bidang studi matematika saja, bukan bidang yang lainnya. Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Oleh karena itu, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung satu saja komponen dominan yang mengukur prestasi subjek. Pada teori respons butir, hubungan antara kemampuan peserta dan skor tes yang dicapai dinyatakan dengan kurva yang tidak linear. Pada Gambar 2 disajikan ilustrasi suatu distribusi kondisional di suatu bagian level kemampuan pada subpopulasi peserta tes. Di sepanjang garis regresi, terdapat sebaran skor tes. Variabilitas kesalahan pengukuran skor tes mungkin terjadi. Jika distribusi bervariasi lintas beberapa subpopulasi, maka tes tidak hanya mengukur kemampuan tunggal saja (Hambleton & Swaminathan, 1985). Jika faktor-faktor yang mempengaruhi prestasi konstan, maka respons subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain. Kondisi ini disebut dengan independensi lokal. Asumsi independensi lokal ini akan terpenuhi apabila jawaban peserta terhadap suatu butir soal mempengaruhi jawaban peserta terhadap terhadap butir soal yang lain. Tes untuk memenuhi asumsi independensi lokal dapat dilakukan dengan membuktikan bahwa peluang dari jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal. Menurut Hambleton, Swaminathan, & Rogers (1991:10), independensi lokal secara matematis dinyatakan sebagai:

$$P(u_1, u_2, \dots, u_n | \theta) = P(u_1 | \theta) \cdot P(u_2 | \theta) \dots P(u_n | \theta)$$

$$P(u_1, u_2, \dots, u_n | \theta) = \prod_{i=1}^n P(u_i | \theta) \quad (1)$$

Keterangan :

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

n: banyaknya butir tes

$P(u_i | \theta)$ : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab butir ke-i dengan benar.

$P(u_1, u_2, \dots, u_n | \theta)$ : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab butir ke-1 sampai ke-n dengan benar

Invariansi parameter artinya karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal. Kemampuan seseorang tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesulitannya dan parameter butir tes tidak akan berubah hanya karena diujikan pada kelompok peserta tes yang berbeda tingkat kemampuannya. Menurut Hambleton, Swaminathan, & Rogers (1991:18), invariansi parameter kemampuan dapat diselidiki dengan mengajukan dua perangkat tes atau lebih yang memiliki tingkat kesukaran yang berbeda pada sekelompok peserta tes. Invariansi parameter kemampuan akan terbukti jika hasil estimasi kemampuan peserta tes tidak berbeda walaupun tes yang dikerjakan berbeda tingkat kesulitannya. Invariansi parameter butir dapat diselidiki dengan mengujikan tes pada kelompok peserta yang berbeda. Invariansi parameter butir terbukti jika hasil estimasi parameter butir tidak berbeda walaupun diujikan pada kelompok peserta yang berbeda tingkat kemampuannya. Dalam teori respons butir, selain asumsi-asumsi yang telah diuraikan sebelumnya, hal penting yang perlu diperhatikan adalah pemilihan model yang tepat. Pemilihan model yang tepat akan mengungkap keadaan yang sesungguhnya dari data tes sebagai hasil pengukuran. Ada 3 model hubungan antara kemampuan dengan parameter butir, yaitu model 1 parameter (model Rasch), model 2 parameter, dan model 3 parameter. Model Rasch dituliskan sebagai berikut:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad (2)$$

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

$P_i(\theta)$ : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dipilih secara acak dapat menjawab butir i dengan benar.

$\theta$ : tingkat kemampuan subyek (sebagai variabel bebas).

b<sub>i</sub>: indeks kesukaran butir ke-i

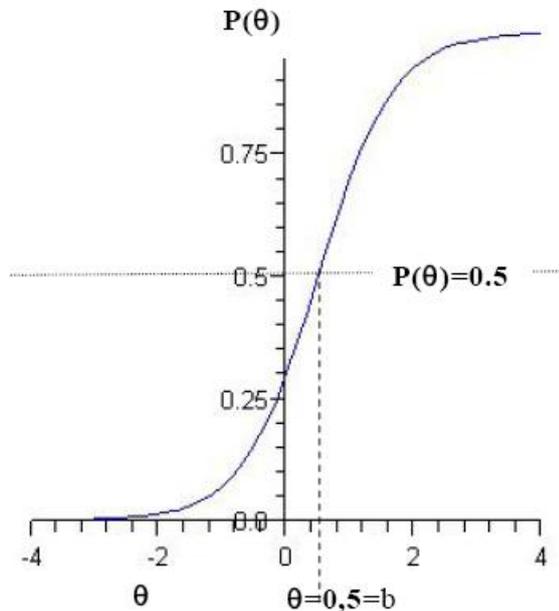
e : bilangan natural yang nilainya mendekati 2,718.

n. : banyaknya butir dalam tes.

Parameter b<sub>i</sub> merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Misalkan suatu butir tes mempunyai parameter b<sub>i</sub> = 0,3, artinya diperlukan kemampuan minimal 0,3 pada skala untuk dapat menjawab benar dengan peluang 50%. Semakin besar nilai parameter b<sub>i</sub>, maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b<sub>i</sub>, maka makin sulit butir soal tersebut. Hubungan peluang menjawab benar  $P_i(\theta)$  dengan tingkat kemampuan peserta ( $\theta$ ) 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 berikut merupakan ilustrasi kurva karakteristik butir untuk model Rasch dengan tingkat kesulitan  $b=0,5$ .



Gambar 1. Kurva Karakteristik Butir untuk Model 1P, dengan  $b=0,5$

Pada model logistik dua parameter, probabilitas peserta tes untuk dapat menjawab benar suatu butir soal ditentukan oleh dua karakteristik butir, yaitu indeks kesukaran butir ( $b_i$ ) dan indeks daya beda butir ( $a_i$ ). Parameter  $a_i$  merupakan indeks daya pembeda yang dimiliki butir ke- $i$ . Pada kurva karakteristik,  $a_i$  proporsional terhadap koefisien arah garis singgung (*slope*) pada titik  $\theta = b$ . Butir soal yang memiliki daya pembeda yang besar mempunyai kurva yang sangat menanjak, sedangkan butir soal yang mempunyai daya pembeda kecil mempunyai kurva yang sangat landai. Secara teoretis, nilai  $a_i$  ini terletak antara  $-\infty$  dan  $+\infty$ . Pada 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 & Swaminathan, 1985:37).

Menurut Hambleton, Swaminathan, & Rogers (1991:15), secara matematis model logistik dua parameter dapat dituliskan sebagai berikut.

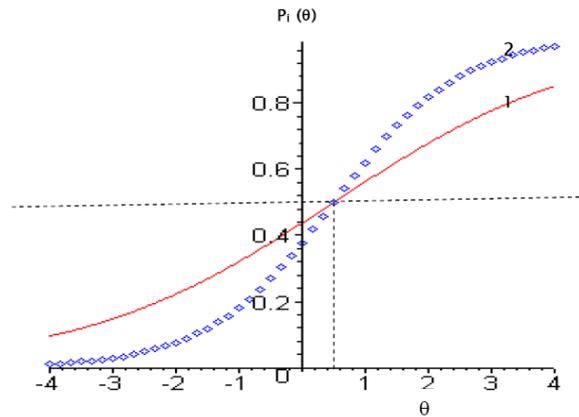
$$P_i(\theta) = \frac{e^{D_i a_i (\theta - b_i)}}{1 + e^{D_i a_i (\theta - b_i)}} \quad (3)$$

dengan  $i : 1, 2, 3, \dots, n$

Keterangan :

- $\theta$  : tingkat kemampuan peserta tes
- $P_i(\theta)$ : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab butir  $i$  dengan benar
- $a_i$  : indeks daya pembeda
- $b_i$  : indeks kesukaran butir ke- $i$
- $e$  : bilangan natural yang nilainya mendekati 2,718
- $n$  : banyaknya butir dalam tes
- $D$  : faktor penskalaan yang harganya 1,7.

Pada gambar 2 disajikan kurva karakteristik butir 1 ( $a=0,5; b=0,5$ ) dan butir 2 ( $a=1; b=0,5$ ). Berdasarkan gambar tersebut, jika indeks daya pembeda butir 1 lebih rendah dibandingkan butir 2, maka akan nampak bahwa kurva karakteristik butir 1 lebih landai dibandingkan butir 2.



Gambar 2. Kurva Karakteristik Butir Model 2P, dengan Butir 1 ( $a=0,5; b=0,5$ ) dan Butir 2 ( $a=1; b=0,5$ )

Sesuai dengan namanya, model logistik tiga parameter ditentukan oleh tiga karakteristik butir yaitu indeks kesukaran butir soal, indeks daya beda butir, dan indeks tebakan semu (*pseudoguessing*). Dengan adanya indeks tebakan semu pada model logistik tiga parameter, memungkinkan subjek yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Secara matematis, model logistik tiga parameter dapat dinyatakan sebagai berikut (Hambleton, & Swaminathan, 1985 : 49; Hambleton, Swaminathan, & Rogers, 1991:17).

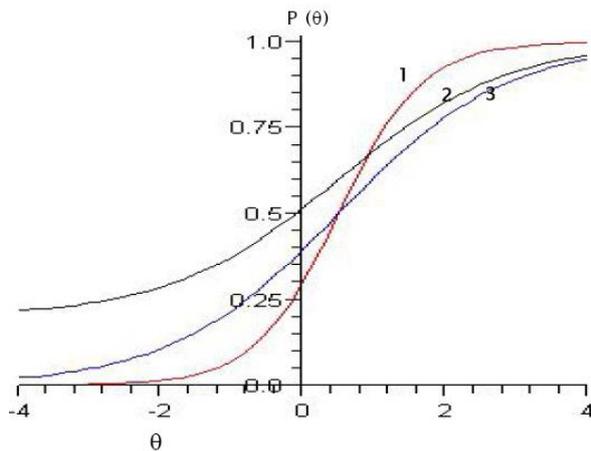
$$P_i(\theta) = c_i + \left\{ (1 - c_i) \frac{e^{D_i a_i (\theta - b_i)}}{1 + e^{D_i a_i (\theta - b_i)}} \right\} \quad (4)$$

Keterangan :

- $\theta$  : tingkat kemampuan peserta tes
- $P_i(\theta)$ : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dapat menjawab butir  $i$  dengan benar
- $a_i$  : indeks daya pembeda
- $b_i$  : indeks kesukaran butir ke- $i$
- $c_i$  : indeks tebakan semu butir ke- $i$
- $e$  : bilangan natural yang nilainya mendekati 2,718
- $n$  : banyaknya butir dalam tes
- $D$  : faktor penskalaan yang harganya 1,7.

Nilai kemampuan peserta ( $\theta$ ) terletak di antara  $-3$  dan  $+3$ , sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai  $b_i$ . Secara teoretis, nilai  $b_i$  terletak di  $\mathbb{R} - \infty$  dan  $+\infty$ . Suatu butir dikatakan baik jika nilai ini berkisar antara  $-2$  dan  $+2$  (Hambleton & 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. Peluang menjawab benar pada saat kemampuan peserta tes sangat rendah dilambangkan dengan  $c_i$ , yang disebut dengan tebakan semu (*pseudoguessing*). Parameter ini merupakan 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. Gambar 3 menyajikan 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 ( $a=0,5, b=0,5, c=0,2$ )

Mencermati gambar tersebut, nampak bahwa pada skala kemampuan peserta tes yang sangat rendah ( $\theta = -4$ ), peluang menjawab benar butir 3 sebesar 0,2, sedangkan pada butir 1 dan butir 2 mendekati 0. 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). Jadi misalkan pada suatu perangkat tes pilihan ganda dengan 4 pilihan untuk setiap butir tesnya, butir ini dikatakan baik jika nilai  $c_i$  tidak lebih dari 0,25. Fungsi informasi butir (*Item Information Functions*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes, dan perbandingan beberapa perangkat tes. Fungsi informasi butir menyatakan kekuatan atau sumbangan 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.

$$I_i(\theta) = \frac{[P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (5)$$

keterangan :

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

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

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

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

$Q_i(\theta)$  : peluang peserta dengan kemampuan  $\theta$  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.

$$I_i(\theta) = \sum_{i=1}^n I_i(\theta) \quad (6)$$

Nilai-nilai indeks parameter butir dan kemampuan peserta merupakan hasil estimasi. Karena merupakan hasil estimasi, maka kebenarannya bersifat probabilitas dan tidak terlepas dengan kesalahan pengukuran. Dalam teori respon butir, kesalahan baku pengukuran (*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  $SEM(\hat{\theta})$ , maka hubungan keduanya, menurut Hambleton, Swaminathan, dan Rogers (1991:94) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad (7)$$

Pada model Rasch 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

$$L(\theta, b; u) = \prod_i \prod_j P_i(\theta_j; b_i)^{u_{ij}} [P_i(\theta_j; b_i)]^{1-u_{ij}} \quad (8)$$

Selanjutnya diestimasi nilai-nilai yang memaksimalkan 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 model logistik Rasch, ada beberapa perangkat lunak yang dapat digunakan, Bigsteps, Rascal, Ascal, Bilog, Xalibrate dan Multilog. Keluaran (*output*) dari program-program ini juga menyediakan hasil pengestimasian parameter peserta tes. Langkah selanjutnya adalah mengetahui kecocokan model dari data yang dianalisis. Uji statistik untuk 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  $\theta$ , yang merupakan dasar untuk mengestimasi nilai  $\theta$ . 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  $\theta_h$ , dengan persamaan :

$$G_j^2 = 2 \sum_{h=1}^{n_h} \left[ r_{hj} \ln \frac{r_{hj}}{N_h P_j(\theta_h)} + (N_h - r_{hj}) \ln \frac{N_h - r_{hj}}{N_h [1 - P(\theta_h)]} \right] \quad (9)$$

dengan  $n_h$  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  $P_j(\theta_h)$  merupakan nilai dari fungsi respons sesuai model untuk butir  $j$  pada  $\theta_h$ , 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  $\theta$ , 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, probabilitas 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). Rumusan penelitian: (1) Bagaimana daya beda butir-butir tes biologi kelas 8 semester gasal? (2) Bagaimana tingkat kesulitan butir-butir tes biologi kelas 8 semester gasal? (3) Bagaimana tebakan tes biologi kelas 8 semester gasal?

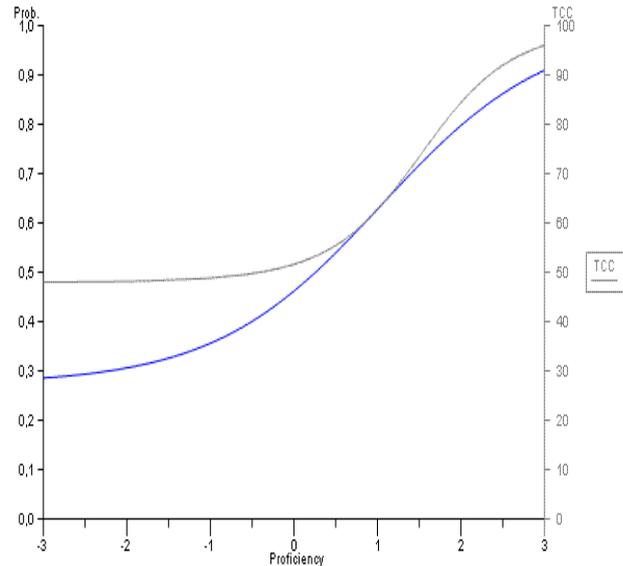
## 2. METODE PENELITIAN

Pengumpulan data dilakukan dengan mengumpulkan lembar jawaban dari 200 responden yang telah mengerjakan tes biologi kelas 8 semester gasal. Setelah semua data terkumpul, maka data dianalisis dengan program komputer, yaitu R programming. Analisis dengan R programming dilakukan untuk mengetahui informasi tentang daya beda, tingkat kesulitan, dan tebakan butir-butir tes biologi kelas 8 semester gasal.

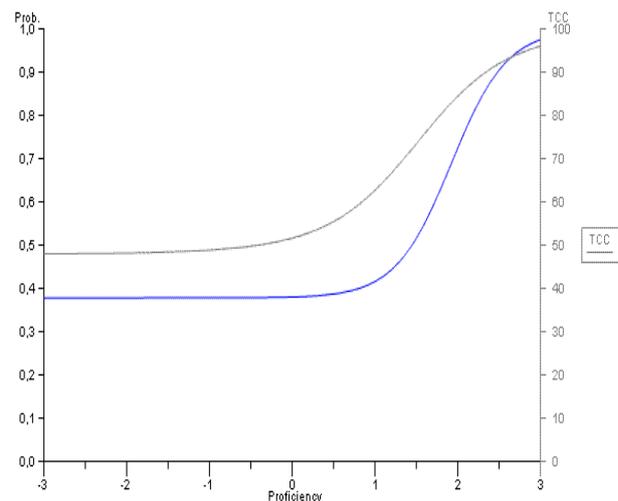
## 3. HASIL PENELITIAN DAN PEMBAHASAN

Daya beda butir terendah adalah 1,002 (butir 8) dan daya beda butir tertinggi adalah 2,969 (butir 38). Tingkat kesulitan butir terendah adalah 1,005 (butir 83 dan 93) dan tingkat kesulitan butir tertinggi adalah 1,992 (butir 45). Tebakan butir terendah adalah 0,018 (butir 80) dan tebakan butir tertinggi adalah 0,989 (butir 68).

Daya beda butir terendah adalah 1,002 (butir 8) dan daya beda butir tertinggi adalah 2,969 (butir 38). Masing-masing butir dilihat kurva karakteristik butir (*item characteristic curve, ICC*) sebagai berikut.



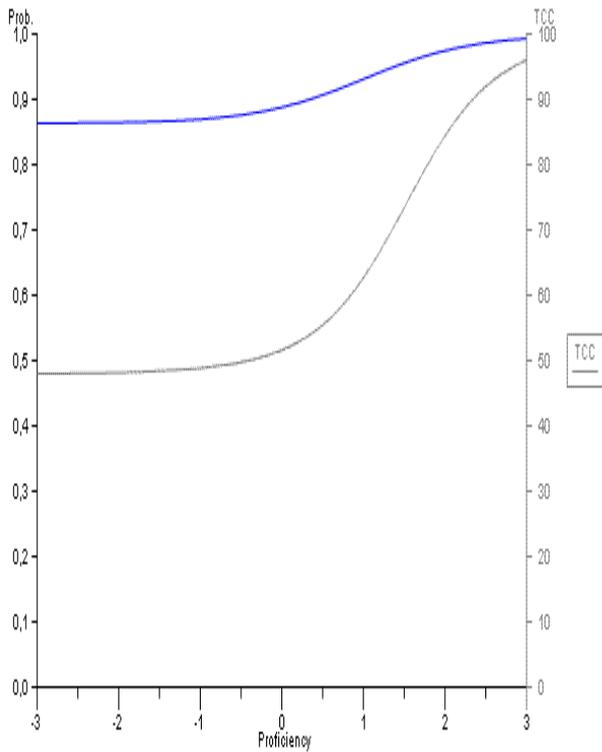
Gambar 4. Kurva Karakteristik Butir 8



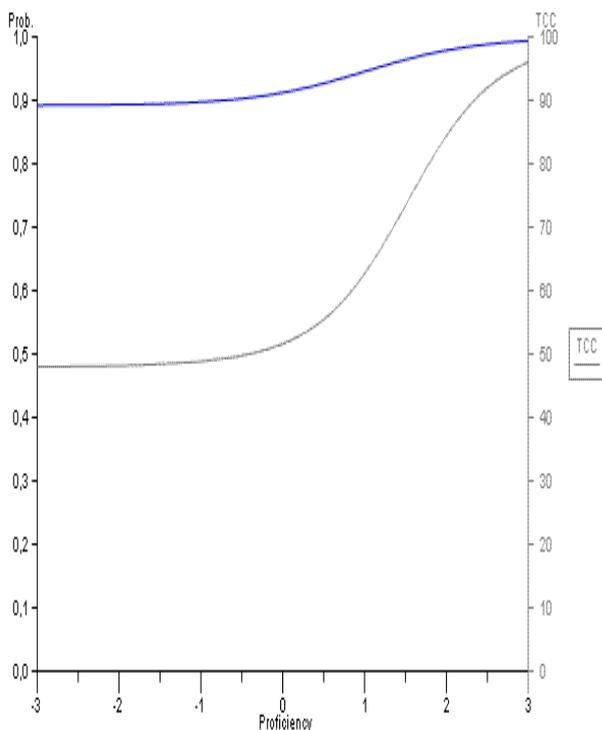
Gambar 5. Kurva Karakteristik Butir 38

Bila dicermati butir 8 mempunyai ICC lebih landai, sedang butir 38 mempunyai ICC lebih yang sangat menanjak, hal ini sesuai dengan teori yang diungkapkan oleh Swaminathan, bahwa butir soal yang memiliki daya pembeda yang besar mempunyai kurva yang sangat menanjak, sedangkan butir soal yang mempunyai daya pembeda kecil mempunyai kurva yang sangat landai (Hambleton & Swaminathan, 1985:37).

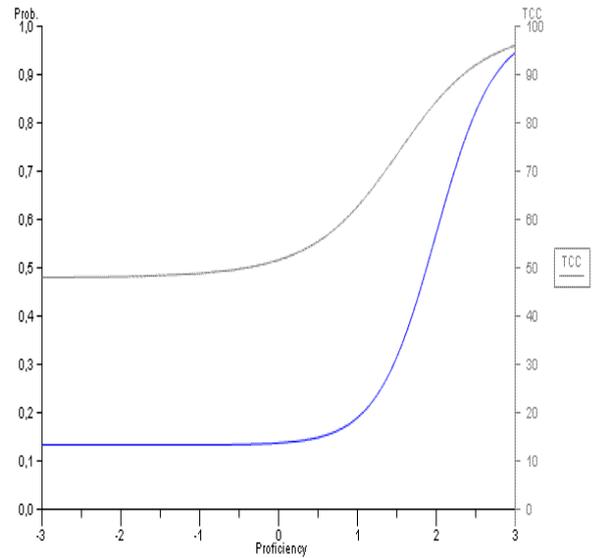
Tingkat kesulitan butir terendah adalah 1,005 (butir 83 dan 93) dan tingkat kesulitan butir tertinggi adalah 1,992 (butir 45). Masing-masing butir dilihat kurva karakteristik butir (*item characteristic curve, ICC*) sebagai berikut.



Gambar 6. Kurva Karakteristik Butir 83



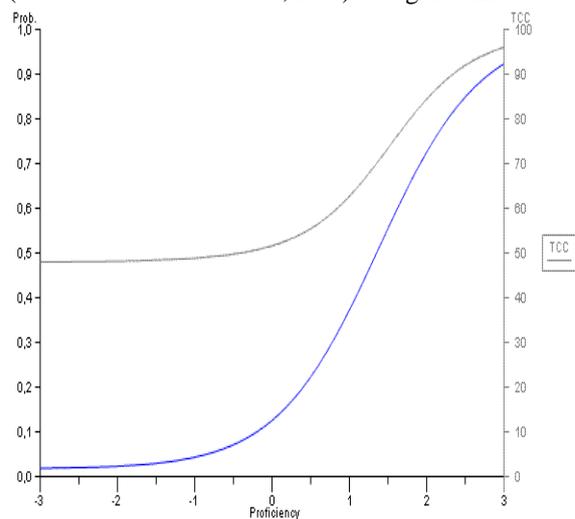
Gambar 7. Kurva Karakteristik Butir 93



Gambar 8. Kurva Karakteristik Butir 45

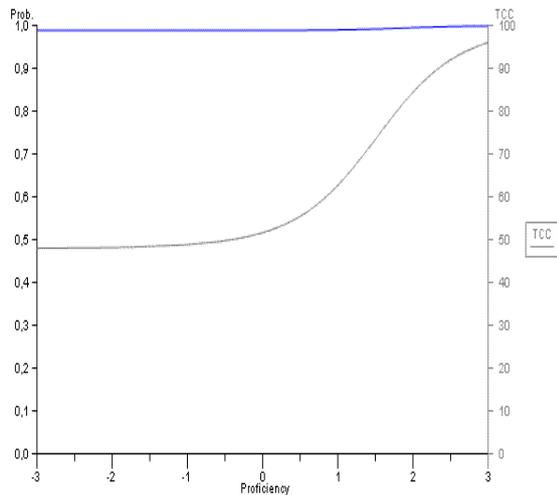
Bila dicermati butir yang mempunyai tingkat kesulitan paling rendah (butir 83, dan butir 93, maka kedua butir ini ICC-nya mendekati  $\theta = -2$ . Sedang butir yang mempunyai tingkat kesulitan yang paling tinggi (butir 45) ICC-nya mendekati  $\theta = +2$ . Fenomena tersebut sesuai dengan yang diungkap oleh Hambleton & Swaminathan, 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 (Hambleton & Swaminathan, 1985:107).

Tebakan butir terendah adalah 0,018 (butir 80) dan tebakan butir tertinggi adalah 0,989 (butir 68). Masing-masing butir dilihat kurva karakteristik butir (*item characteristic curve, ICC*) sebagai berikut.



Gambar 9. Kurva Karakteristik Butir 80





Gambar 10. Kurva Karakteristik Butir 68

Bila dicermati butir yang mempunyai tebakan paling rendah (butir 80), maka ICC-nya mendekati 0,000, sedang butir yang mempunyai tebakan sangat tinggi (butir 68), maka ICC-nya mendekati 1,00. Oleh sebab itu butir 68 perlu didrop karena termasuk butir soal yang tidak baik. Fenomena ini sesuai dengan teori yang diungkap oleh Hullin, bahwa suatu butir tes, nilai  $ci$  berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai  $ci$  tidak lebih dari  $1/k$ , dengan  $k$  banyaknya pilihan. Jadi pada suatu perangkat tes pilihan ganda dengan 4 pilihan untuk setiap butir tesnya, butir dikatakan baik jika nilai  $ci$  tidak lebih dari 0,25 (Hullin, 1983:36).

#### 4. SIMPULAN

Simpulan yang dapat diambil adalah: (1) Daya beda butir-butir tes biologi kelas 8 semester gasal berkisar 1,002 sampai 2,969. Daya beda butir yang paling rendah adalah butir 8 dan daya beda butir yang paling tinggi adalah butir 38. (2) Tingkat kesulitan butir tes biologi kelas 8 semester gasal berkisar 1,005 sampai 1,992. Tingkat kesulitan yang paling rendah butir 83 dan 93 dan tingkat kesulitan yang paling tinggi adalah butir 45. (3) Tebakan butir-butir tes biologi kelas 8 semester gasal berkisar 0,018 sampai 0,989. Tebakan yang paling rendah adalah butir 80 dan tebakan yang paling tinggi adalah butir 68.

#### 5. UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Kemenristekdikti yang telah memberikan dana untuk terlaksananya penelitian ini, sehingga peneliti dapat melakukan penelitian dengan lancar dan dapat membuat artikel yang dapat disampaikan dalam acara seminar nasional ini.

#### 6. DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ : Prentice Hall.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (Ed.6). New York: Harper & Row.
- Ebel, R. L. (1979). *Essential of educational measurement*. New Jersey: Prentice Hall Inc.
- Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of educational measurement*. Englewood Cliffs: Prentice- Hall, Inc.
- Gronlund, N.E. (1976). *Measurement and evaluation in teaching*. New York: Macmillan Publishing Co.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6<sup>th</sup>ed). New York: Macmillan.
- 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). *Keberfungsian butir diferensial pada perangkat tes seleksi masuk SMP*. Unpublished Master Tesis, Program Studi Penelitian dan Evaluasi Pendidikan sekolah Pascasarjana Universitas Negeri Yogyakarta.
- 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.
- Kerlinger, F.N. (1986). *Asas-asas penelitian behavioral* (Terjemahan L.R. Simatupang). Yogyakarta: Gajahmada University Press.
- Keeves, J.P. dan Alagumalai, S. (1999). *New approaches to measurement*. Dalam
- Linn, R.L. (1989). *Educational measurement*. (3<sup>rd</sup> ed.). New York: Macmillan Publishing Company.
- 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*. Mooreville : Scientific Software Inc.
- Syaifudin Azwar. (2000). *Reliabilitas dan validitas* (Edisi 4). Yogyakarta: Pustaka Pelajar.
- Suwarto. (2013). *Pengembangan tes diagnostik dalam pembelajaran*. Yogyakarta: Pustaka Relajar

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

**Penanya:** Nura Syifa

**Pertanyaan:**

Tes yang dibuat akan digunakan untuk mengukur kemampuan peserta apa untuk mengukur kompetensi?

**Jawaban:**

Mengukur kompetensi atau mengukur kemampuan pada prinsipnya sama, sehingga tes biologi yang saya buat dapat digunakan untuk mengukur kedua kemampuan pada peserta tes.

