

# KARAKTERISTIK TES BIOLOGI KELAS 7 SEMESTER GASAL

*Suwarto*

Program Studi Pendidikan Biologi-FKIP  
Universitas Veteran Bangun Nusantara Sukoharjo  
Jl. Sujono Humardani No. 1 Jombor Sukoharjo  
e-mail: suwartowarto@yahoo.com

## ABSTRAK

*Penelitian ini bertujuan untuk: (1) memaparkan tingkat kesulitan butir tes biologi kelas 7 semester gasal, (2) memaparkan daya beda butir-butir tes biologi kelas 7 semester gasal, dan (3) mengetahui reliabilitas tes biologi kelas 7 semester gasal. Tes biologi kelas 7 semester gasal terdiri dari 100 butir soal dan diujicobakan pada 200 responden. Teknik analisis data yang digunakan adalah program iteman versi 3.00. Hasil analisis menunjukkan: (1) Tingkat kesulitan butir-butir tes biologi kelas 7 semester gasal berkisar 0,025 sampai 0,775. Tingkat kesulitan butir yang paling sulit adalah butir 30 dan tingkat kesulitan butir yang paling mudah adalah butir 75. Perbandingan persentase butir yang mudah: butir yang sedang: butir yang sulit adalah 26%:29%:45%, (2) Daya beda butir-butir tes biologi kelas 7 semester gasal berkisar -0,226 sampai 1,000. Daya beda yang paling rendah adalah butir 47 dan daya beda butir yang paling tinggi adalah butir 65 dan butir 66. Butir soal yang mempunyai daya beda butir jelek ada 18 butir. Butir soal yang mempunyai daya beda butir cukup ada 23 butir. Butir soal yang mempunyai daya beda butir baik ada 22 butir. Butir soal yang mempunyai daya beda butir sangat baik ada 37 butir, (3) Reliabilitas tes biologi kelas 7 semester gasal adalah 0,839.*

**Kata kunci:** analisis butir soal, tes biologi, tingkat kesulitan

## ABSTRACT

*This study deals with Biology test item of the odd semester grade 7. Its aims are as follows: (1) to describe the difficulties of Biology test items, (2) to describe the standard differentials, and (3) to determine the reliability of the tests. The test consists of 100 items and tested on 200 respondents. Data analysis technique used is the iteman program version 3:00. Results of the analysis showed: (1) Level of difficulty of the test ranges from 0.025 to 0.775. The most difficult item is item 30 and the easiest one is item 75. Comparison of percentage of the difficulties of high, mid, and low 26%: 29%: 45%, (2) the standard differential ranges from -0.226 to 1.000. Item 47 has the lowest discrimination index and items 64 and 65 have the highest. Item 18 has the poorest discrimination index, item 23 the marginal one, item 22 the reasonably good one, and item 37 the very good one, (3) the reliability of the test is 0,839.*

**Keywords:** item test analysis, Biologi test, item difficulty

## PENDAHULUAN

Ujian semester di sekolah lanjutan tingkat pertama (SLTP) selalu 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) mengungkapkan bahwa tes adalah seperangkat pertanyaan yang dibuat untuk diberikan kepada siswa dengan syarat-syarat tertentu. Dengan kata lain, tes adalah prosedur yang sistematis untuk mengobservasi tingkah laku. Gronlund & Lin, Cronbach, dan Rusli mendefinisikan tes yang dipandang dari sudut prosedur pengukuran. Prosedur pengukuran dilakukan secara sistematis untuk mendapatkan informasi tingkah laku siswa.

Uno, Sofyan, dan Candiasa (2001:62) menjelaskan bahwa tes merupakan seperangkat rangsangan (stimuli) yang diberikan kepada seseorang dengan maksud untuk mendapatkan jawaban-jawaban yang dapat dijadikan dasar bagi penetapan skor angka. Djemari (2004:71) tes merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah. Tes diartikan juga sebagai sejumlah pertanyaan yang membutuhkan jawaban atau sejumlah pernyataan yang harus diberikan tanggapan, dengan tujuan mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang dikenai tes. Menurut Suryabrata (1987:27) tes adalah pertanyaan yang harus dijawab atau perintah yang harus dijalankan dan atas jawaban tersebut orang dapat mengambil kesimpulan dengan cara membandingkan dengan suatu standar atau peserta tes lain. Ahli pengukuran yang lain yaitu Ebel menjelaskan kegunaan tes untuk keperluan yang lebih luas. Ebel (1979:380) mengungkapkan bahwa *a test is any kind of device or procedure for measuring ability, achievement, interest, and other traits*.

Berdasarkan pernyataan di atas, tes merupakan alat bantu untuk mengukur perilaku yang terdiri atas pertanyaan yang berguna sebagai penyaring informasi perilaku yang dimiliki individu dengan bantuan sebuah skala numerik atau sebuah kategori sistem. Pengukuran dapat didefinisikan sebagai *the process by which information about the attributes or characteristics of thing are determined and differentiated* (Oriondo, 1998:2). Guilford mendefinisikan pengukuran dengan *assigning numbers to, or quantifying, things according to a set of rules* (Griffin dan Nix, 1991:3). Pengukuran dinyatakan sebagai proses penetapan angka terhadap individu atau karakteristiknya menurut aturan tertentu (Ebel & Frisbie, 1986:14). Allen & Yen (1979:2) mendefinisikan pengukuran sebagai penetapan angka dengan cara yang sistematis untuk menyatakan keadaan individu. Keadaan individu ini bisa berupa kemampuan kognitif, afektif, dan psikomotor. Bahkan, akhir-akhir ini dikembangkan kemampuan emosi, yaitu kemampuan yang menentukan kesuksesan seseorang dalam melaksanakan pekerjaan (Djemari, 2000:1). Sax (1980:9) mengungkapkan bahwa pengukuran adalah kegiatan yang menyangkut pemberian angka-angka terhadap atribut atau ciri-ciri seseorang, benda, atau kejadian sesuai dengan aturan atau rumus-rumus. Pengukuran prestasi belajar siswa juga dilakukan dengan kuantitas berdasarkan aturan-aturan tertentu. Pengukuran memfokuskan pada sifat spesifik dan menggunakan angka untuk menentukan tingkat penguasaan terhadap sifat itu. Pengukuran merupakan proses penetapan angka-angka atau kategori tertentu terhadap hasil belajar siswa untuk menggambarkan kualitas hasil belajar siswa.

*The Task Group on Assessment and Testing* (Griffin & Nix, 1991:3) mendeskripsikan asesmen sebagai semua cara yang digunakan untuk menilai unjuk kerja individu atau kelompok. Popham (1996:3) mendefinisikan asesmen dalam konteks pendidikan sebagai sebuah usaha secara

formal untuk menentukan status siswa berkenaan dengan berbagai kepentingan pendidikan. Boyer & Ewel mendefinisikan *asesment* sebagai “*processes that provide information about individual students, about curricula or programs, about institutions, or about entire systems of institutions*” (Stark & Thomas, 1994:46). Dalam konteks pendidikan, asesmen berkaitan dengan semua proses pendidikan, seperti karakteristik peserta didik, karakteristik metode mengajar, kurikulum, fasilitas, dan administrasi.

Nitko (1996:33) mengungkapkan bahwa penilaian mengarah kepada proses memperoleh informasi yang relevan untuk tujuan-tujuan dalam rangka mengambil keputusan, bukan alat memperoleh informasi. Dengan demikian, penilaian harus belajar merupakan aktivitas yang direncanakan dan dilaksanakan oleh guru untuk mendapatkan data dan informasi guna mengambil keputusan. Data dan informasi yang diperoleh merupakan umpan-balik tentang tingkat keberhasilan proses pembelajaran dan informasi tentang kemampuan belajar yang dicapai siswa. Untuk memperoleh data dan informasi yang akurat, diperlukan teknik penilaian yang tepat. Ketepatan penentuan alat penilaian dapat berpengaruh terhadap hasil penilaian dan hasil penilaian akan mempengaruhi keputusan yang akan diambil.

Tujuan tes yang penting adalah untuk: (a) mengetahui tingkat kemampuan siswa, (b) mengukur pertumbuhan dan perkembangan siswa, (c) mendiagnosis kesulitan belajar siswa, (d) mengetahui hasil pengajaran, (e) mengetahui hasil belajar, (f) mengetahui pencapaian kurikulum, (g) mendorong siswa belajar, dan (h) mendorong guru agar mengajar yang lebih baik (Djemari, 2004:72).

Sering kali tes digunakan untuk beberapa tujuan, namun tidak akan memiliki keefektifan yang sama untuk semua tujuan. Ditinjau dari tujuannya, ada empat macam tes yang digunakan di banyak lembaga pendidikan, yaitu (a) tes penempatan, (b) tes diagnostik, (c) tes formatif, dan (d) tes sumatif. Pengujian berbasis kemampuan dasar pada umumnya menggunakan tes diagnostik, formatif, dan sumatif. Adapun tes biologi kelas 7 semester gasal bertujuan untuk mengetahui tingkat kemampuan siswa dalam memahami konsep biologi.

Tes yang baik harus terdiri atas butir-butir soal yang baik. Pada tes pilihan ganda, butir soal yang baik harus memiliki tingkat kesulitan yang memadai, daya pembeda yang baik, dan berfungsinya pengecoh. Oleh karena itu, dalam mengembangkan tes pilihan ganda harus diperhatikan tingkat kesulitan, daya pembeda, dan berfungsinya pengecoh.

Tingkat kesulitan menunjuk kepada perbandingan antara banyaknya peserta tes yang menjawab benar dengan banyaknya seluruh peserta tes (Suwanto, 2013:105). Daya pembeda menunjuk kepada selisih proporsi yang menjawab benar pada kelompok atas dan proporsi yang menjawab benar pada kelompok bawah (Suwanto, 2013:109). Pada perkembangannya, daya pembeda suatu butir didefinisikan sebagai korelasi antara skor butir tersebut dengan skor total (McDonald, 1999:231). Berfungsinya pengecoh menunjuk kepada seberapa banyak peserta yang memiliki pengecoh tersebut. Suatu butir soal pilihan ganda dikatakan memenuhi persyaratan apabila besarnya tingkat kesulitan berkisar antara 0,30 dan 0,80, besarnya daya pembeda 0,30 atau lebih, dan pengecoh dipilih oleh paling sedikit 5% dari seluruh peserta tes (Djemari, 2002:116).

Validitas merupakan salah satu aspek yang penting yang harus dimiliki setiap tes. Allen & Yen (1979:95) mengemukakan *a test has validity if it measures what it purporst to measures* artinya suatu tes dikatakan valid jika tes tersebut mengukur yang seharusnya diukur. Allen & Yen (1979:95) juga menyatakan *test is valid if its scores are highly related to examinees future performance on the job*. Artinya, suatu tes dikatakan valid apabila skor tes tersebut sangat berkaitan dengan menguji penampilan masa depan dalam pekerjaan seseorang.

Menurut Djemari (2004:25), validitas ialah ukuran seberapa cermat suatu tes melakukan fungsi ukurnya. Test hanya dapat melakukan fungsinya dengan cermat kalau ada “sesuatu”

yang diukurnya. Jadi, untuk dikatakan valid, tes harus mengukur sesuatu dan melakukannya dengan cermat. Dua ahli pengukuran tersebut dapat disimpulkan bahwa validitas merupakan karakteristik yang harus dimiliki oleh suatu tes. Suatu tes dikatakan valid jika mengukur apa yang mau diukur. Suatu tes digunakan untuk tujuan tertentu sehingga validitas tes tersebut berbeda-beda.

Allen & Yen (1979:72) *there are several ways of defining and interpreting test reliability*. Artinya, ada beberapa cara penentuan dan interpretasi reliabilitas tes. Alat ukur yang reliabel adalah suatu alat ukur yang mantap tidak berubah-ubah hasil pengukurannya dan dapat diandalkan. Penggunaan alat ukur tersebut secara berulang-ulang akan memberikan hasil yang konsisten. Sifat reliabel dari sebuah alat ukur berkenaan dengan kemampuan alat ukur tersebut memberikan hasil yang konsisten. Suhr (2003:6) *development of a measurement instrument is a complex process. Reliability assesses the accuracy and precision of the instrument*. Pengembangan suatu instrumen pengukuran adalah proses yang kompleks. Reliabilitas menaksir ketelitian dan ketepatan dari suatu instrumen.

Jahja Umar (1991) menjelaskan reliabilitas mempunyai dua konsistensi, yaitu konsistensi internal dan konsistensi eksternal. Konsistensi internal adalah tingkat sejauh mana butir soal itu homogen, baik dari segi tingkat kesukaran maupun bentuk soalnya. Konsistensi eksternal adalah sejauh mana skor yang dihasilkan tetap sama sepanjang orang yang diukur kemampuannya belum berubah. Hasil pengukuran terhadap seseorang siswa diharapkan akan sama apabila pengukuran itu diulang.

Dengan demikian, reliabilitas adalah tingkat ketepatan, kejelasan, atau kemantapan. Suatu alat ukur disebut mempunyai reliabilitas tinggi atau dapat dipercaya jika alat ukur itu mantap, dalam pengertian bahwa alat ukur tersebut stabil, dapat diandalkan dan dapat digunakan untuk meramalkan. Batas bawah koefisien reliabilitas yang digunakan untuk suatu tes yang baik, yaitu sebesar 0,700 (Linn, 1989:106; Suwanto, 2013:176).

Rumusan penelitian ini dirumuskan sebagai berikut. (1) Bagaimana tingkat kesulitan butir-butir tes biologi kelas 7 semester gasa? (2) Bagaimana daya beda butir-butir tes biologi kelas 7 semester gasal? (3) Bagaimana reliabilitas tes biologi kelas 7 semester gasal?

## METODE PENELITIAN

Data diperoleh dengan mengumpulkan lembar jawaban dari 200 responden yang telah mengerjakan tes biologi kelas 7 semester gasal. Setelah semua data terkumpul, data dianalisis dengan program komputer, yaitu program ITEMAN versi 3.00. Analisis dengan program ITEMAN dilakukan untuk mengetahui informasi tentang tingkat kesukaran butir (*proportion correct = p*), daya beda butir (*biserial* atau *point biserial*), alternatif yang menyatakan banyaknya alternatif dalam tiap butir, *proportion endorsing* yang menyatakan proporsi untuk tiap-tiap alternatif jawaban, *biserial* dan *point biserial* untuk tiap-tiap alternatif jawaban. Sejumlah kriteria analisis butir soal antara lain tercantum dalam tabel 1 berikut ini.

**Tabel 1. Beberapa Kriteria pada Analisis Butir Soal**

Karakteristik	Kategori	Kriteria
Tingkat Kesukaran	Mudah	$P > 0,700$
	Sedang	$0,300 \leq p \leq 0,700$
	Sukar	$P < 0,300$

Daya Beda	Jelek	$Biser \leq 0,200$
	Cukup	$0,200 < biser \leq 0,400$
	Baik	$0,410 < biser \leq 0,700$
	Sangat baik	$0,710 < biser \leq 1,000$
Distraktor	Tidak berfungsi	Dipilih $< 5\%$ peserta
	Berfungsi	Dipilih $\geq 5\%$ peserta
Kunci	Perlu dicek	$Biser \ \& \ point \ biser \ key < biser \ \& \ point \ biser \ distraktor$
	Baik	$Biser \ \& \ point \ biser \ key > biser \ \& \ point \ biser \ distraktor$
Reliabilitas	Tidak reliabel	$< 0,700$
	Reliabel	$> 0,700$

## HASIL DAN PEMBAHASAN

Tingkat kesulitan butir (*Prop. Correct*) terendah adalah 0,025 (butir 30) dan tingkat kesulitan butir tertinggi adalah 0,775 (butir 75). Dari data tersebut dapat diperoleh pengertian bahwa butir yang paling sulit adalah butir 30, sedangkan butir yang paling mudah adalah butir 75. Adapun kategori tingkat kesulitan butir tes biologi kelas 7 semester gasal selengkapannya dapat dirangkum sebagai berikut.

**Tabel 2. Kategori Tingkat Kesulitan Butir**

Kategori	Nomor Butir	Jumlah
Mudah ( $Prop. Correct > 0,700$ )	75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 94, 95, 96, 97, 98, 99, 100	26
Sedang ( $0,300 \leq Prop. Correct \leq 0,700$ )	2, 3, 4, 5, 6, 7, 9, 10, 12, 14, 18, 22, 29, 41, 43, 44, 46, 47, 48, 49, 51, 53, 59, 63, 64, 68, 70, 71, 73	29
Sulit ( $Prop. Correct < 0,300$ )	1, 8, 11, 13, 15, 16, 17, 19, 20, 21, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 45, 50, 52, 54, 55, 56, 57, 58, 60, 61, 62, 65, 66, 67, 69, 72, 74	45
	Jumlah	100

Dari Tabel 2 dapat diketahui bahwa tingkat kesulitan butir yang termasuk kategori mudah ada 26 butir. Tingkat kesulitan butir yang tergolong mudah ada  $26/100 \times 100\% = 26\%$ . Tingkat kesulitan butir yang termasuk kategori sedang ada 29 butir. Tingkat kesulitan butir yang tergolong sedang ada  $29/100 \times 100\% = 29\%$ . Tingkat kesulitan butir yang termasuk kategori sulit ada 45 butir. Tingkat kesulitan butir yang tergolong sulit ada  $45/100 \times 100\% = 45\%$ . Dari persentase tingkat kesulitan butir dari masing-masing kategori dapat diketahui bahwa sebagian besar tingkat kesulitan butir adalah kategori sulit.

Daya beda butir (*Biser*) terendah adalah -0,226 (butir 47) dan daya beda butir tertinggi adalah 1,000 (butir 65, dan butir 66). Butir yang mempunyai daya beda negatif harus didrop, yaitu butir 32, 41, 47, 50, dan 53. Adapun kategori daya beda butir tes biologi kelas 7 semester

gasal selengkapnya dapat dilihat pada Tabel 3.

**Tabel 3. Kategori Daya Beda Butir**

Kategori	Nomor Butir	Jumlah
Jelek (Biser $\leq$ 0,200)	20, 21, 23, 24, 26, 27, 30, 32, 33, 37, 38, 41, 42, 45, 47, 50, 53, 64	18
Cukup (0,200 < Biser $\leq$ 0,400)	3, 4, 8, 10, 14, 15, 18, 19, 22, 25, 28, 29, 35, 36, 39, 43, 44, 48, 59, 60, 63, 71, 75	23
Baik (0,400 < Biser $\leq$ 0,700)	1, 2, 5, 6, 7, 9, 11, 12, 13, 16, 34, 46, 49, 51, 52, 54, 56, 58, 68, 69, 70, 73	22
Sangat Baik (0,700 < Biser < 1,000)	17, 31, 40, 55, 57, 61, 62, 65, 66, 67, 72, 74, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 94, 95, 96, 97, 98, 99, 100	37
	Jumlah	100

Dari Tabel 3 dapat diketahui bahwa daya beda butir yang termasuk kategori jelek ada 18 butir. Daya beda butir yang tergolong jelek ada  $18/100 \times 100\% = 18\%$ . Butir yang tergolong jelek ini seharusnya tidak boleh digunakan (Biser  $\leq$  0,200 harus didrop). Daya beda butir yang termasuk kategori cukup ada 23 butir. Daya beda butir yang tergolong cukup ada  $23/100 \times 100\% = 23\%$ . Daya beda butir yang tergolong cukup ini masih perlu direvisi. Daya beda butir yang termasuk kategori baik ada 22 butir. Daya beda butir yang tergolong baik ada  $22/100 \times 100\% = 22\%$ . Daya beda butir yang termasuk kategori sangat baik ada 37 butir. Daya beda butir yang tergolong sangat baik ada  $37/100 \times 100\% = 37\%$ . Dari persentase daya beda butir dari masing-masing kategori dapat diketahui bahwa sebagian besar daya beda butir adalah sangat baik.

*Option* yang kurang berfungsi (*Prop.Endorsing* < 0,050 atau kurang dari 5%) adalah: *option* E butir 2, *option* D butir 4, *option* C dan D butir 10, *option* E butir 17, *option* A butir 18, *option* E butir 20, *option* B butir 30, *option* E butir 36, *option* A butir 39, *option* E butir 43, *option* E butir 48, *option* A, B, dan D butir 61, *option* A butir 62, *option* E butir 63, *option* A butir 64, *option* A butir 66, *option* D butir 68, *option* D butir 69, *option* C dan E butir 70, *option* C dan E butir 71, *option* E butir 72, *option* B butir 73, *option* C butir 74, *option* B dan D butir 75. Dengan demikian ada 30 *options* yang perlu direvisi.

*Key* (kunci) jawaban perlu diteliti kembali kebenarannya karena kurang diminati oleh responden dan responden cenderung memilih distraktornya. Pada *output* analisis dengan program iteman akan muncul CHECK THE KEY menunjukkan perlu diteliti kembali kunci jawaban yang telah dibuat dan mengapa responden cenderung tertarik memilih distraktornya. Butir-butir yang perlu dicek kunci jawabannya adalah butir: 20, 21, 22, 25, 26, 28, 30, 32, 33, 37, 38, 41, 42, 45, 47, 50, 53, 63, dan 64. Sehingga ada 19 butir yang perlu diteliti kembali kunci jawabannya.

Reliabilitas tes biologi kelas 7 semester gasal adalah 0,839. Reliabilitas ini tergolong baik karena reliabilitas suatu instrumen dapat digunakan apabila minimal 0,700. Tes yang baik mempunyai butir-butir soal dengan persentase 25% mudah, 50% sedang dan 25% sulit. Bila diperhatikan, hasil analisis tes biologi kelas 7 semester gasal dengan program ITEMAN versi 3.00 dari sudut pandang tingkat kesulitan butir ternyata persentase butir mudah: persentase butir sedang: persentase butir sulit adalah 26% : 29% : 45%. Hal ini menunjukkan komposisi persentase butir mudah, butir sedang, dan butir sulit pada tes yang baik tidak terpenuhi.

Daya beda butir soal yang dapat digunakan sebagai tes yang baik adalah daya beda butir  $\geq$  0,400. Adapun dari hasil analisis tes biologi kelas 7 semester gasal ternyata daya beda yang

baik 22 butir dan daya beda yang sangat baik 37 butir. Dengan demikian, butir-butir soal yang telah memenuhi syarat minimal dari sudut pandang daya beda butir hanya ada 59 butir. Butir-butir yang belum memenuhi persyaratan dari sudut pandang daya beda butir sebanyak 41 butir.

*Option* yang masih harus direvisi juga masih banyak yaitu ada 30 *option*. *Option* yang memenuhi persyaratan *option* yang baik adalah *option* yang dipilih oleh responden minimal 5% atau 0,050. Demikian juga masih banyak butir yang perlu ditinjau ulang kunci jawabannya, yaitu sebanyak 19 butir.

Reliabilitas tes biologi kelas 7 semester gasal tergolong baik yaitu 0,839 sudah lebih besar dari 0,700. Reliabilitas instrumen dapat digunakan jika reliabilitas instrumen minimal 0,700 (Linn, 1989: 106; Suwanto, 2013: 176).

## SIMPULAN

Simpulan yang dapat diambil dari penelitian ini dapat dijabarkan sebagai berikut. (1) Tingkat kesulitan butir-butir tes biologi kelas 7 semester gasal berkisar dari 0,025 sampai 0,775. Tingkat kesulitan butir yang paling sulit adalah butir 30 dan tingkat kesulitan butir yang paling mudah adalah butir 75. Perbandingan persentase butir yang mudah: butir yang sedang: butir yang sulit adalah 26% : 29% : 45%. (2) Daya beda butir-butir tes biologi kelas 7 semester gasal berkisar dari -0,226 sampai 1,000. Daya beda yang paling rendah adalah butir 47 dan daya beda butir yang paling tinggi adalah butir 65 dan butir 66. Butir soal yang mempunyai daya beda butir jelek ada 18 butir. Butir soal yang mempunyai daya beda butir cukup ada 23 butir. Butir soal yang mempunyai daya beda butir baik ada 22 butir. Butir soal yang mempunyai daya beda butir sangat baik ada 37 butir. (3) Reliabilitas tes biologi kelas 7 semester gasal adalah 0,839.

## DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey: Brooks/Cole Publishing Company.
- Cronbach, L. J. (1970). *Essentials of Psychological Testing* (Ed.6). New York: Harper & Row.
- Djemari Mardapi. (2000). "Pengembangan dan Validasi Instrumen Akreditasi." Makalah Disampaikan pada Pelatihan Akreditasi Madrasah Bagi Pejabat Kanwil Depag, di Yogyakarta.
- Djemari Mardapi. (2002). "Bukti Kesahihan dan Keandalan Alat Ukur: Tanggapan Atas Artikel "Tes Keterampilan Olah Raga Judo bagi Mahasiswa". *Jurnal Kependidikan*. No.1 tahun XXXII. Lembaga Penelitian UNY.
- Djemari Mardapi. (2004). *Penyusunan Tes Hasil Belajar*. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Ebel, R. L. (1979). *Essential ff Educational Measurement*. New Jersey: Prentice Hall Inc.
- Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of Educational Measurement*. Englewood Cliffs: Prentice- Hall, Inc.
- Griffin, P. & Nix, P. (1991). *Educational Assessment and Reporting*. Sydney: Harcourt Brace Javanovich, Publisher.

- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching (6<sup>th</sup>ed)*. New York: Macmillan.
- Linn, R.L. (1989). *Educational Measurement*. (3<sup>rd</sup> ed.). New York: Macmillan Publishing Company.
- Nitko, A.J. (1996). *Curriculum-based Assessment Workshop Papers*. Jakarta: Directorate of General of Primary and Secondary Education.
- Oriondo, L. L. & Antonio, E. M.D. (1998). *Evaluating Educational Outcomes (Test, Measurement and Evaluation)*. Manila: Rex Book Store
- Popham, W. J. (1995). *Classroom Assessment*. Boston: Allyn and Bacon.
- Rusli, R.S. (1988). *Tes dan Pengukuran dalam Pendidikan*. Jakarta: Depdikbud.
- Stark, J.S. & Thomas, A. (1994). *Assessment and Program Evaluation*. Needham Heights: Simon & Schuster Custom Publishing.
- Suhr, D. (2003). *Reliability, Exploratory & Confirmatory Factor Analysis for the Scale of Athletic Priorities*. Diambil pada tanggal 2 Agustus 2006, dari: <http://www2.sas.com/proceedings/sugi28/274-28.pdf>
- Suryabrata, S. (1987). *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali.
- Suwarto. (2013). *Pengembangan Tes Diagnostik dalam Pembelajaran*. Yogyakarta: Pustaka Pelajar
- Uno, H.B., Sofyan, H., & Candiasa, I.M. (2001). *Pengembangan Instrumen untuk Penelitian*. Jakarta: Delima Press.