



# PROSIDING

## Seminar Nasional MIPA 2016

Naskah diseminarkan pada 5 November 2016 dan dipublikasikan pada  
<http://conf.unnes.ac.id/index.php/mipa/mipa2016/schedConf/presentations>



## Tingkat Kesulitan, Daya Beda, dan Reliabilitas Tes Biologi Kelas 7 Semester Genap

**Suwarto**

Program Studi Pendidikan Biologi-FKIP  
Universitas Veteran Bangun Nusantara Sukoharjo  
e-mail: suwartowarto@yahoo.com

### Abstrak

Penelitian ini bertujuan untuk: (1) memaparkan tingkat kesulitan butir tes biologi kelas 7 semester genap, (2) memaparkan daya beda butir tes biologi kelas 7 semester genap, (3) mengetahui reliabilitas tes biologi kelas 7 semester genap. Tes biologi kelas 7 semester genap terdiri dari 100 butir soal dan diujicobakan pada 300 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 genap berkisar 0,024 sampai 0,776. Tingkat kesulitan butir yang paling sulit adalah butir 11 dan tingkat kesulitan butir yang paling mudah adalah butir 50. Perbandingan persentase butir yang mudah: butir yang sedang: butir yang sulit adalah 25%:50%:25%, (2) Daya beda butir-butir tes biologi kelas 7 semester genap berkisar -0,226 sampai 1,000. Daya beda yang paling rendah adalah butir 1 dan daya beda butir yang paling tinggi adalah butir 4 dan butir 10. Butir soal yang mempunyai daya beda butir jelek ada 10 butir. Butir soal yang mempunyai daya beda butir cukup ada 30 butir. Butir soal yang mempunyai daya beda butir baik ada 35 butir. Butir soal yang mempunyai daya beda butir sangat baik ada 25 butir, (3) Reliabilitas tes biologi kelas 7 semester genap adalah 0,833.

**Kata-kata Kunci:** analisis butir soal

### Abstract

*This study deals with Biology test item of the even 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 describe the reliability of the tests. The test consists of 100 items and tested on 300 respondents. Data analysis technique used is The item program version 3.00. Result of the analysis showed: (1) Level of difficulty of the test ranges from 0,024 to 0,776. The most difficult item is item 11 and the easiest one is item 50. Comparison of percentage of the difficulties of high, mid, and low 25% : 50% : 25%, (2) The standard difficultial range from -0,226 to 1,000. Item 1 has the lowes discrimination index and item 4 and 10 have the highest. Item 10 the poorest discrimination index, item 30 the marginal one, item 35 the reseonably good one, and item 25 the very good one, (3) The reliability of the test is 0,833.*

**Keywords:** item test analysis

## PENDAHULUAN

Ilmu Pengetahuan Alam (IPA) berkaitan dengan cara mencari tahu tentang alam secara sistematis, sehingga IPA bukan hanya penguasaan kumpulan pengetahuan yang berupa fakta-fakta, konsep-konsep, atau prinsip-prinsip saja tetapi juga merupakan suatu proses penemuan. Pendidikan IPA diharapkan dapat menjadi wahana bagi peserta didik

untuk mempelajari diri sendiri dan alam sekitar, serta prospek pengembangan lebih lanjut dalam menerapkannya di dalam kehidupan sehari-hari. Proses pembelajarannya menekankan pada pemberian pengalaman langsung untuk mengembangkan kompetensi agar menjelajahi dan memahami alam sekitar secara ilmiah. Pendidikan IPA diarahkan untuk inkuiri dan berbuat sehingga dapat membantu peserta didik untuk memperoleh pemahaman yang lebih mendalam tentang alam sekitar. Standar Kompetensi (SK) dan Kompetensi Dasar (KD) IPA di SMP/MTs merupakan standar minimum yang secara nasional harus dicapai oleh peserta didik dan menjadi acuan dalam pengembangan kurikulum di setiap satuan pendidikan. Pencapaian SK dan KD didasarkan pada pemberdayaan peserta didik untuk membangun kemampuan, bekerja ilmiah, dan pengetahuan sendiri yang difasilitasi oleh guru.

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) *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) 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*. S. Eko Putro Widoyoko (2012) mendefinisikan bahwa tes merupakan salah satu alat untuk melakukan pengukuran, yaitu alat untuk mengumpulkan informasi karakteristik suatu objek. Gronlund & Lin, Cronbach, dan S. Eko Putro Widoyoko sejalan dalam mendefinisikan tes. Mereka mendefinisikan tes yang dipandang dari sudut prosedur pengukuran dalam rangka mengumpulkan informasi. Prosedur pengukuran dilakukan secara sistematis untuk mendapatkan informasi tingkah laku siswa (Suwarto, 2016).

Uno, Sofyan, & Candiasa (2001) 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) 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. Suryabrata (1987) 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) *a test is any kind of device or procedure for measuring ability, achievement, interest, and other traits* (Suwarto, 2016).

Berdasarkan pernyataan di atas tes merupakan alat bantu untuk mengukur perilaku, 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). Guilford mendefinisikan pengukuran dengan *assigning numbers to, or quantifying, things according to a set of rules* (Griffin dan Nix, 1991). Pengukuran dinyatakan sebagai proses penetapan angka terhadap individu atau karakteristiknya menurut aturan tertentu (Ebel & Frisbie, 1986). Allen & Yen (1979) 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). Sax (1980) 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 (Suwarto, 2016).

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).

Seringkali tes digunakan untuk beberapa tujuan, namun tidak akan memiliki keefektifan yang sama untuk semua tujuan. Ditinjau dari tujuannya, ada empat macam tes yang banyak digunakan di lembaga pendidikan, yaitu: (a) tes penempatan, (b) tes diagnostic, (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 genap 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 berfungsi pengecoh. Oleh karena itu, dalam mengembangkan tes pilihan ganda harus diperhatikan tingkat kesulitan, daya pembeda, dan berfungsi pengecoh (Suwarto, 2016).

Tingkat kesulitan menunjuk kepada perbandingan antara banyaknya peserta tes yang menjawab benar dengan banyaknya seluruh peserta tes (Suwarto, 2013). Daya pembeda menunjuk kepada selisih proporsi yang menjawab benar pada kelompok atas dan proporsi yang menjawab benar pada kelompok bawah (Suwarto, 2013). Pada perkembangannya, daya pembeda suatu butir didefinisikan sebagai korelasi antara skor butir tersebut dengan skor total (McDonald, 1999). 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).

Validitas merupakan salah satu aspek yang penting yang harus dimiliki setiap tes. Allen & Yen (1979) 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) 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.

Tabel 1 Kriteria 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	$Biser < 0,200$
	Cukup	$0,200 \leq biser < 0,400$
	Baik	$0,400 \leq biser < 0,700$
	Sangat baik	$0,700 \leq biser \leq 1,000$
Distraktor	Tidak berfungsi	Dipilih $< 5\%$ peserta
	Berfungsi	Dipilih $\geq 5\%$ peserta
Kunci	Perlu dicek	$Biser \& \text{ point biser key} < biser \& \text{ point biser distraktor}$
	Baik	$Biser \& \text{ point biser key} > biser \& \text{ point biser distraktor}$
Reliabilitas	Tidak reliabel	$< 0,700$
	Reliabel	$> 0,700$

(Suwarto, 2016)

Jahja Umar (1999) 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 bahwa reliabilitas adalah tingkat ketepatan, keajegan, 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; Suwarto, 2013; Suwarto, 2016).

Rumusan penelitian: (1) Bagaimana tingkat kesulitan butir-butir tes biologi kelas 7 semester genap? (2) Bagaimana daya beda butir-butir tes biologi kelas 7 semester genap? (3) Bagaimana reliabilitas tes biologi kelas 7 semester genap.

## METODE

Ujicoba tes dilakukan di SMP se Solo Raya. Pengumpulan data dilakukan dengan mengumpulkan lembar jawaban dari 300 responden yang telah mengerjakan tes biologi kelas 7 semester genap. Setelah semua data terkumpul, maka 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

Tabel 2. Kategori Tingkat Kesulitan Butir

Kategori	Nomor Butir	Jumlah
Mudah ( $P > 0,700$ )	2, 10, 14, 16, 23, 25, 28, 32, 33, 35, 39, 41, 44, 50, 51, 56, 67, 71, 77, 82, 83, 88, 90, 94, 99	25
Sedang ( $0,300 \leq P \leq 0,700$ )	1, 3, 4, 7, 9, 12, 15, 17, 20, 21, 22, 24, 26, 27, 29, 30, 34, 36, 38, 40, 42, 45, 46, 47, 49, 52, 54, 57, 60, 63, 64, 65, 66, 68, 69, 70, 72, 74, 76, 78, 79, 80, 81, 84, 85, 89, 91, 92, 93, 100	50
Sulit ( $P < 0,300$ )	5, 6, 8, 11, 13, 18, 19, 31, 37, 43, 48, 53, 55, 58, 59, 61, 62, 73, 75, 86, 87, 95, 96, 97, 98	25
Jumlah		100

Tabel 3. Kategori Daya Beda Butir

Kategori	Nomor Butir	Jumlah
Jelek ( $D \leq 0,200$ )	1, 5, 8, 21, 30, 31, 37, 61, 70, 91	10
Cukup ( $0,200 < D \leq 0,400$ )	3, 6, 9, 13, 16, 19, 23, 26, 29, 33, 36, 39, 43, 46, 49, 53, 56, 59, 63, 66, 69, 73, 76, 79, 83, 86, 88, 93, 96, 98	30
Baik ( $0,400 < D \leq 0,700$ )	2, 7, 11, 12, 15, 17, 18, 22, 25, 27, 28, 32, 35, 38, 41, 42, 45, 47, 48, 51, 52, 55, 58, 62, 65, 68, 71, 72, 75, 78, 81, 82, 85, 92, 95	35
Sangat Baik ( $0,700 < D < 1,000$ )	4, 10, 14, 20, 24, 34, 40, 44, 50, 54, 57, 60, 64, 67, 74, 77, 80, 84, 87, 89, 90, 94, 97, 99, 100	25
Jumlah		100

tiap-tiap alternatif jawaban. Sejumlah kriteria analisis butir soal antara lain tercantum dalam Tabel 1.

## HASIL DAN PEMBAHASAN

Tingkat kesulitan butir (*Prop. Correct*) yang selanjutnya disingkat dengan *p* terendah adalah 0,024 (butir 11) dan tingkat kesulitan butir tertinggi adalah 0,776 (butir 50). Dari data tersebut dapat diperoleh pengertian bahwa butir yang paling sulit adalah butir 11 sedangkan butir yang paling mudah adalah butir 50. Adapun kategori tingkat kesulitan butir tes biologi kelas 7 semester genap selengkapnya dapat dirangkum sebagai berikut.

Dari Tabel 2 dapat diketahui bahwa tingkat kesulitan butir yang termasuk kategori mudah ada 25 butir. Tingkat kesulitan butir yang tergolong mudah ada  $25/100 \times 100\% = 25\%$ . Tingkat kesulitan butir yang termasuk kategori sedang ada 50 butir. Tingkat kesulitan butir yang tergolong sedang ada  $50/100 \times 100\% = 50\%$ . Tingkat kesulitan butir yang termasuk kategori sulit ada 25 butir. Tingkat kesulitan butir yang tergolong sulit ada  $25/100 \times 100\% = 25\%$ . Dari persentase tingkat kesulitan butir dari masing-masing kategori dapat diketahui bahwa sebagian besar tingkat kesulitan butir adalah kategori sedang.

Daya beda butir (*Biser*) yang selanjutnya disingkat *D* terendah adalah -0,226 (butir 1) dan daya beda butir tertinggi adalah 1,000 (butir 4, dan butir 10). Butir yang mempunyai daya beda negatif harus didrop, yaitu butir 1, 5, 8, 21, dan 30. Adapun kategori daya beda butir tes biologi kelas 7 semester genap selengkapnya dapat dilihat pada Tabel 3.

Dari Tabel 3 dapat diketahui bahwa daya beda butir yang termasuk kategori jelek ada 10 butir. Daya beda butir yang tergolong jelek ada  $10/100 \times 100\% = 10\%$ . Butir yang tergolong jelek ini seharusnya tidak boleh digunakan ( $Biser \leq 0,200$  harus didrop). Daya beda butir yang termasuk kategori cukup ada 30 butir. Daya beda butir yang tergolong cukup ada  $30/100 \times 100\% = 30\%$ . Daya beda butir yang tergolong cukup ini masih perlu direvisi. Daya beda butir yang termasuk kategori baik ada 35 butir. Daya beda butir yang tergolong baik ada  $35/100 \times 100\% = 35\%$ . Daya beda butir yang termasuk kategori sangat baik ada 25 butir. Daya beda butir yang tergolong sangat baik ada  $25/100 \times 100\% = 25\%$ . Dari persentase daya beda butir dari masing-masing kategori dapat diketahui bahwa sebagian besar daya beda butir adalah baik.

*Option* yang kurang berfungsi (*Prop.Endorsing* < 0,050 atau kurang dari 5%) adalah: *option* D butir 3, *option* C butir 4, *option* C dan D butir 11, *option* D butir 20, *option* A butir 20, *option* D butir 20, *option* B butir 30, *option* D butir 36, *option* A butir 40, *option* D butir 45, *option* D butir 48, *option* A, B, dan D butir 63, *option* A butir 62, *option* D butir 63, *option* A butir 64, *option* A butir 66, *option* D butir 68, *option* D butir 69, *option* C dan D butir 70, *option* C dan D butir 71, *option* D butir 72, *option* B butir 73, *option* C butir 74, *option* B dan D butir 75. Dengan demikian ada 30 *option* yang perlu direvisi.

*Key* (kunci) jawaban yang 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: 19, 20, 23, 25, 27, 29, 30, 31, 40, dan 65. Sehingga ada 10 butir yang perlu diteliti kembali kunci jawabannya. Reliabilitas tes biologi kelas 7 semester genap adalah 0,833. 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 genap dengan program ITEMAN versi 3.00 dari sudut pandang tingkat kesulitan butir ternyata persentase butir mudah: persentase butir sedang: persentase butir sulit adalah 25%:50%:25%. Hal ini menunjukkan komposisi persentase butir mudah, butir sedang, dan butir sulit pada tes yang baik sudah terpenuhi.

Daya beda butir soal yang dapat digunakan sebagai tes yang baik adalah daya beda butir  $\geq 0,400$ . Sedangkan dari hasil analisis tes biologi kelas 7 semester genap ternyata daya beda yang baik 35 butir, dan daya beda yang sangat baik 25 butir. Dengan demikian maka butir-butir soal yang telah memenuhi syarat minimal dari sudut pandang daya beda butir hanya ada 60 butir. Butir-butir yang belum memenuhi persyaratan dari sudut pandang daya beda butir sebanyak 40 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-butir yang perlu ditinjau ulang kunci jawabannya, yaitu sebanyak 10 butir.

Reliabilitas tes biologi kelas 7 semester genap tergolong baik yaitu 0,833 sudah lebih besar dari 0,700. Reliabilitas instrumen dapat digunakan jika reliabilitas instrumen minimal 0,700 (Linn, 1989: 106; Suwarto, 2013: 176; Suwarto, 2016).

## SIMPULAN

Kesimpulan yang dapat diambil adalah: (1) Tingkat kesulitan butir-butir tes biologi kelas 7 semester genap berkisar dari 0,024 sampai 0,776. Tingkat kesulitan butir yang paling sulit adalah butir 11 dan tingkat kesulitan butir yang paling mudah adalah butir 50. Perbandingan persentase butir yang mudah: butir yang sedang: butir yang sulit adalah 25%:50%:25%. (2) Daya beda butir-butir tes biologi kelas 7 semester genap berkisar dari -0,226 sampai 1,000. Daya beda yang paling rendah adalah butir 1 dan daya beda butir yang paling tinggi adalah butir 4 dan butir 10. Butir soal yang mempunyai daya beda butir jelek ada 10 butir. Butir soal yang mempunyai daya beda butir cukup ada 30 butir. Butir soal yang mempunyai daya beda butir baik ada 35 butir. Butir soal yang mempunyai daya beda butir sangat baik ada 25 butir. (3) Reliabilitas tes biologi kelas 7 semester genap adalah 0,833.

## 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 of 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). Newyork: Macmillan.
- Jahja Umar (1999). Item banking. Dalam G. N. Master & J. P. Keeves. *Advances in measurement in educational research and assessment* (pp. 207-219). Amsterdam-Lausanne-New York-Oxford-Shannon-Singapore-Tokyo: Pergamon
- Linn, R.L. (1989). *Educational measurement*. (3<sup>rd</sup> ed.). New York: Macmillan Publishing Company.
- McDonald, P.R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlmbaum Associates, Inc.
- 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.
- Stark, J.S. & Thomas, A. (1994). *Assessment and program evaluation*. Needham Heights: Simon & Schuster Custom Publishing.
- S. Eko Putro Widoyoko (2012). *Evaluasi program pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Suryabrata, S. (1987). *Pengembangan tes hasil belajar*. Jakarta: Rajawali.
- Suwarto. (2013). *Pengembangan tes diagnostik dalam pembelajaran*. Yogyakarta: Pustaka Pelajar
- Suwarto. (2016). Karakteristik tes biologi kelas 7 semester gasal. *Jurnal Penelitian Humaniora* Vol. 17, No. 1, Februari 2016. Hal: 1-8.
- Uno, H.B., Sofyan, H., & Candiasa, I.M. (2001). *Pengembangan instrumen untuk penelitian*. Jakarta: Delima Press.