

Karakteristik Tes Biologi Kelas 7 Semester Gasal

The Biology Test Characteristic of 7th 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: e-mail: suwartowarto@yahoo.com

Abstract: The goals of the research: (1) To know the difficulty degree of the Biology test items for class VII by the period of the odd term, (2) To know the interval of the Biology test items for class VII by the period of the odd term, (3) To know the reliability of of the Biology test for class VII by the period of the odd term. The Biology test for class VII by the period of the odd term consists of 100 items tested toward 100 respondents. The item programming version 3.00 is used in the analysis. The result of the analysis shows that: (1) The difficulty degree of the Biology test items for class VII by the period of the odd term is at the interval of 0,025 up to 0,775. There are 30 items which are difficultly tested and 75 items which are easily tested. The comparison of the easy tested items: the medium tested items : the difficult tested items is 26% : 29% : 45%, (2) The interval of the Biology test items for class VII by the period of the odd term shows between -0,226 up to 1,000. The lowest interval belongs to the item numbered 47 and the highest interval belongs to the items numbered 65 and 66. There are found 18 items which belong in bad qualification, 23 items which belong in medium qualification, 22 items which belong in good qualification and 37 items which belong in the best qualification, (3) The reliability of of the Biology test for class VII by the period of the odd term is 0,839.

Keywords: The test characteristic

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.

Uno, Sofyan, & Candiasa (2001:62) 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. 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) *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, 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 mendefinisi 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) 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 asesmen 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) 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 maka 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).

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 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 diukur. 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) 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, 1989:106; Suwanto, 2013:176).

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

2. METODE PENELITIAN

Pengumpulan data dilakukan dengan mengumpulkan lembar jawaban dari 200 responden yang telah mengerjakan tes biologi kelas 7 semester gasal. 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 tiap-tiap alternatif jawaban.

3. HASIL PENELITIAN DAN PEMBAHASAN

Tingkat kesulitan butir (*Prop. Correct*) yang selanjutnya disingkat dengan p 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 selengkapnya dapat dirangkum sebagai berikut.

Tabel 1. Kategori Tingkat Kesulitan Butir

Kategori	Nomor Butir	Jumlah
Mudah ($P > 0,700$)	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 ($0,300 \leq P \leq 0,700$)	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 ($P < 0,300$)	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 1 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*) yang selanjutnya disingkat D 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 2.

Tabel 2. Kategori Daya Beda Butir

Kategori	Nomor Butir	Jumlah
Jelek ($D \leq 0,200$)	20, 21, 23, 24, 26, 27, 30, 32, 33, 37, 38, 41, 42, 45, 47, 50, 53, 64	18
Cukup ($0,200 < D \leq 0,400$)	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 ($0,400 < D \leq 0,700$)	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 ($0,700 < D < 1,000$)	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,	37

Kategori	Nomor Butir	Jumlah
	94, 95, 96, 97, 98, 99, 100	
	Jumlah	100

Dari Tabel 2 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 *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: 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 $\geq 0,400$. Sedangkan 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 maka 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-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).

4. SIMPULAN

Simpulan yang dapat diambil adalah: (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.

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.
- 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 (6thed)*. Newyork: Macmillan.
- Linn, R.L. (1989). *Educational measurement*. (3rd 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, Retrieved from <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.

