

**PENYUSUNAN ALAT PENILAIAN HASIL BELAJAR
DAN ANALISIS BUTIR SOALNYA**

Oleh: Ali Muhson, M.Pd.

BAGIAN I. PENILAIAN KELAS

A. Pengertian Penilaian Kelas

Penilaian kelas adalah suatu bentuk kegiatan guru yang terkait dengan pengambilan keputusan tentang pencapaian kompetensi atau hasil belajar peserta didik yang mengikuti proses pembelajaran tertentu. Untuk itu, diperlukan data sebagai informasi yang diandalkan sebagai dasar pengambilan keputusan. Dalam hal ini, keputusan berhubungan dengan sudah atau belum berhasilnya peserta didik dalam mencapai suatu kompetensi. Jadi, penilaian kelas merupakan salah satu pilar dalam pelaksanaan Kurikulum Tingkat Satuan Pendidikan (KTSP) yang berbasis kompetensi.

Data yang diperoleh guru selama pembelajaran berlangsung dapat diaring dan dikumpulkan melalui prosedur dan alat penilaian yang sesuai dengan kompetensi atau hasil belajar yang akan dinilai. Oleh sebab itu, penilaian kelas lebih merupakan proses pengumpulan dan penggunaan informasi oleh guru untuk memberikan keputusan, dalam hal ini nilai terhadap hasil belajar peserta didik berdasarkan tahapan belajarnya. Dari proses ini, diperoleh potret/profil kemampuan peserta didik dalam mencapai sejumlah Standar Kompetensi dan kompetensi dasar yang tercantum dalam kurikulum.

Penilaian kelas merupakan suatu proses yang dilakukan melalui langkah-langkah perencanaan, penyusunan alat penilaian, pengumpulan informasi melalui sejumlah bukti yang menunjukkan pencapaian hasil belajar peserta didik, pengolahan, dan penggunaan informasi tentang hasil belajar peserta didik. Penilaian kelas dilaksanakan melalui berbagai cara, seperti unjuk kerja (performance), penilaian sikap, penilaian tertulis (paper and pencil test), penilaian proyek, penilaian produk, penilaian melalui kumpulan hasil kerja/karya peserta didik (portfolio), , dan

penilaian diri.

Penilaian hasil belajar baik formal maupun informal diadakan dalam suasana yang menyenangkan, sehingga memungkinkan peserta didik menunjukkan apa yang dipahami dan mampu dikerjakannya. Hasil belajar seorang peserta didik tidak dianjurkan untuk dibandingkan dengan peserta didik lainnya, tetapi dengan hasil yang dimiliki peserta didik tersebut sebelumnya. Dengan demikian peserta didik tidak merasa dihakimi oleh guru tetapi dibantu untuk mencapai apa yang diharapkan.

B. Manfaat Penilaian Kelas

Manfaat penilaian kelas antara lain sebagai berikut

1. Untuk memberikan umpan balik bagi peserta didik agar mengetahui kekuatan dan kelemahannya dalam proses pencapaian kompetensi.
2. Untuk memantau kemajuan dan mendiagnosis kesulitan belajar yang dialami peserta didik sehingga dapat dilakukan pengayaan dan remedial.
3. Untuk umpan balik bagi guru dalam memperbaiki metode, pendekatan, kegiatan, dan sumber belajar yang digunakan.
4. Untuk masukan bagi guru guna merancang kegiatan belajar.
5. Untuk memberikan informasi kepada orangtua dan komite sekolah tentang efektivitas pendidikan.
6. Untuk memberi umpan balik bagi pengambil kebijakan (Diknas Daerah) dalam mempertimbangkan konsep penilaian kelas yang baik untuk digunakan

C. Fungsi Penilaian Kelas

Penilaian kelas memiliki fungsi sebagai berikut

1. Memberikan informasi sejauhmana seorang peserta didik telah menguasai suatu kompetensi.
2. Mengevaluasi hasil belajar peserta didik dalam rangka membantu peserta didik memahami dirinya, membuat keputusan tentang langkah berikutnya, baik untuk

- pemilihan program, pengembangan kepribadian maupun untuk penjurusan (sebagai bimbingan).
3. Menemukan kesulitan belajar peserta didik kemungkinan prestasi yang bisa dikembangkan peserta didik dan sebagai alat diagnosis yang membantu guru menentukan apakah seseorang perlu mengikuti remedial atau pengayaan.
 4. Menemukan kelemahan dan kekurangan proses pembelajaran yang sedang berlangsung guna perbaikan proses pembelajaran berikutnya.
 5. Sebagai kontrol bagi guru dan sekolah tentang kemajuan perkembangan peserta didik.

D. Rambu-rambu Penilaian Kelas

1. Kriteria Penilaian Kelas

a. Validitas

Validitas berarti menilai apa yang seharusnya dinilai dengan menggunakan alat yang sesuai untuk mengukur kompetensi. Dalam menyusun soal sebagai alat penilaian perlu memperhatikan kompetensi yang diukur, dan menggunakan bahasa yang tidak mengandung makna ganda. Misal, dalam pelajaran bahasa Indonesia, guru ingin menilai kompetensi berbicara. Bentuk penilaian valid jika menggunakan tes lisan. Jika menggunakan tes tertulis penilaian tidak valid.

b. Reliabilitas

Reliabilitas berkaitan dengan konsistensi (kejegan) hasil penilaian. Penilaian yang reliable (ajeg) memungkinkan perbandingan yang reliable dan menjamin konsistensi. Misal, guru menilai suatu proyek, penilaian akan reliabel jika hasil yang diperoleh itu cenderung sama bila proyek itu dilakukan lagi dengan kondisi yang relatif sama. Untuk menjamin penilaian yang reliabel petunjuk pelaksanaan proyek dan penSkorannya harus jelas.

c. Terfokus pada kompetensi

Dalam pelaksanaan kurikulum tingkat satuan pendidikan yang berbasis

kompetensi, penilaian harus terfokus pada pencapaian kompetensi (rangkaiannya kemampuan), bukan hanya pada penguasaan materi (pengetahuan).

d. Keseluruhan/Komprehensif

Penilaian harus menyeluruh dengan menggunakan beragam cara dan alat untuk menilai beragam kompetensi atau kemampuan peserta didik, sehingga tergambar profil kemampuan peserta didik.

e. Objektivitas

Penilaian harus dilaksanakan secara obyektif. Untuk itu, penilaian harus adil, terencana, berkesinambungan, dan menerapkan kriteria yang jelas dalam pemberian Skor.

f. Mendidik

Penilaian dilakukan untuk memperbaiki proses pembelajaran bagi guru dan meningkatkan kualitas belajar bagi peserta didik.

2. Prinsip Penilaian Kelas

Dalam melaksanakan penilaian, guru seyogianya:

- a. Memandang penilaian dan kegiatan pembelajaran secara terpadu, sehingga penilaian berjalan bersama-sama dengan proses pembelajaran.
- b. Mengembangkan tugas-tugas penilaian yang bermakna, terkait langsung dengan kehidupan nyata.
- c. Mengembangkan strategi yang mendorong dan memperkuat penilaian sebagai cermin diri.
- d. Melakukan berbagai strategi penilaian di dalam program pembelajaran untuk menyediakan berbagai jenis informasi tentang hasil belajar peserta didik.
- e. Mempertimbangkan berbagai kebutuhan khusus peserta didik.
- f. Mengembangkan dan menyediakan sistem pencatatan yang bervariasi dalam pengamatan kegiatan belajar peserta didik.
- g. Menggunakan cara dan alat penilaian yang bervariasi. Penilaian kelas dapat

dilakukan dengan cara tertulis, lisan, produk, portofolio, unjuk kerja, proyek, dan pengamatan partisipasi peserta didik dalam proses pembelajaran sehari-hari sesuai dengan kompetensi dasar yang harus dikuasai.

- h. Melakukan Penilaian kelas secara berkesinambungan terhadap semua Stándar Kompetensi dan Kompetensi Dasar untuk memantau proses, kemajuan, dan perbaikan hasil dalam bentuk ulangan harian, ulangan tengah semester, ulangan akhir semester, dan ulangan kenaikan kelas.
- i. Mengadakan ulangan harian bila sudah menyelesaikan satu atau beberapa indikator. Dengan demikian tidak perlu menunggu menyelesaikan 1 KD, karena ruang lingkupnya besar. Pelaksanaan ulangan harian dapat dilakukan dengan penilaian tertulis, penilaian lisan, penilaian unjuk kerja, atau bentuk lain yang sesuai dengan karakteristik materi atau kompetensi yang dinilai. Ulangan tengah semester dilakukan bila telah menyelesaikan beberapa kompetensi dasar dipertengahan semester, sedangkan ulangan akhir semester dilakukan setelah menyelesaikan semua kompetensi dasar semester bersangkutan. Ulangan kenaikan kelas dilakukan pada akhir semester genap dengan menilai semua kompetensi dasar semester ganjil dan genap, dengan penekanan pada kompetensi dasar semester genap. Guru menetapkan tingkat pencapaian kompetensi peserta didik berdasarkan hasil belajarnya pada kurun waktu tertentu (akhir semester atau akhir tahun).

Agar penilaian objektif, guru harus berupaya secara optimal untuk (1) memanfaatkan berbagai bukti hasil kerja peserta didik dan tingkah laku dari sejumlah penilaian, (2) membuat keputusan yang adil tentang penguasaan kompetensi peserta didik dengan mempertimbangkan hasil kerja (karya) mereka.

3. Penilaian pada Pendidikan Sistem Ganda

Kurikulum SMK berbasis kompetensi (competency-based curriculum), berbasis luas dan mendasar (broad-based curriculum), dan menggunakan pendekatan pengembangan kecakapan hidup (life Standar Kompetensi skills

development approach). Pelaksanaan pembelajaran di SMK menggunakan model penyelenggaraan Pendidikan Sistem Ganda (PSG), yaitu pembelajaran dilaksanakan di sekolah dan di dunia usaha/dunia industri (Du/Di), dengan menggunakan kurikulum yang disusun oleh sekolah sesuai dengan tuntutan dunia kerja atau kurikulum yang disusun bersama antara Du/Di dengan sekolah. Karena itu, dimungkinkan melibatkan Du/Di sebagai penilai.

Penilaian kompetensi pada PSG (uji kompetensi) melibatkan pihak sekolah, Pusat, Asosiasi Profesi/LSP (Lembaga Sertifikasi Profesi), dan pihak lain terutama DU/DI. Idealnya, lembaga yang menyelenggarakan uji kompetensi ini independen; yakni lembaga yang tidak dapat diintervensi oleh unsur atau lembaga lain.

4. Penilaian Hasil Belajar Masing-masing Kelompok Mata Pelajaran

- a. Penilaian hasil belajar kelompok mata pelajaran agama dan akhlak mulia serta kelompok mata pelajaran kewarganegaraan dan kepribadian dilakukan melalui:
 - 1). Pengamatan terhadap perubahan perilaku dan sikap untuk menilai perkembangan afeksi dan kepribadian peserta didik
 - 2). Ujian, ulangan, dan/atau penugasan untuk mengukur aspek kognitif peserta didik
- b. Penilaian hasil belajar kelompok mata pelajaran ilmu pengetahuan dan teknologi diukur melalui ulangan, penugasan, dan/atau bentuk lain yang sesuai dengan karakteristik materi yang dinilai
- c. Penilaian hasil belajar kelompok mata pelajaran estetika dilakukan melalui pengamatan terhadap perubahan perilaku dan sikap untuk menilai perkembangan afeksi dan ekspresi psikomotorik peserta didik.
- d. Penilaian hasil belajar kelompok mata pelajaran jasmani, olahraga, dan kesehatan dilakukan melalui:
 - 1). Pengamatan terhadap perubahan perilaku dan sikap untuk menilai

- perkembangan psikomotorik dan afeksi peserta didik; dan
- 2). Ulangan, dan/atau penugasan untuk mengukur aspek kognitif peserta didik.

E. Ranah Penilaian Berdasarkan Kurikulum Berbasis Kompetensi

Kurikulum berbasis kompetensi tidak semata-mata meningkatkan pengetahuan peserta didik, tetapi kompetensi secara utuh yang merefleksikan pengetahuan, keterampilan, dan sikap sesuai karakteristik masing-masing mata pelajaran. Dengan kata lain, kurikulum tersebut menuntut proses pembelajaran di sekolah berorientasi pada penguasaan kompetensi-kompetensi yang telah ditentukan.

Kurikulum ini memuat sejumlah kompetensi untuk setiap matadiklat. Satu standar kompetensi terdiri dari beberapa kompetensi dasar. Pada kurikulum tingkat satuan pendidikan, satu kompetensi dasar dikembangkan menjadi beberapa indikator pencapaian kompetensi. Indikator-indikator tersebut menjadi acuan dalam merancang dan melaksanakan penilaian pembelajaran.

BAGIAN II. ANALISIS BUTIR SOAL

A. Pendahuluan

Untuk mendapatkan instrumen berkualitas tinggi, selain dilakukan analisis secara teori (telaah butir berdasarkan aspek isi, konstruksi, dan bahasa) perlu juga dilakukan analisis butir secara empirik. Secara garis besar, analisis butir secara empirik ini dapat dibedakan menjadi dua, yaitu dengan pendekatan teori tes klasik (*Classical Test Theory*) dan teori respons butir (*Item Response Theory, IRT*).

Teori tes klasik atau disebut teori skor murni klasik (Allen & Yen, 1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Jika dituliskan dengan pernyataan matematis, maka kalimat tersebut menjadi

$$X = T + E$$

dengan :

X : skor amatan,

T : skor sebenarnya,

E : skor kesalahan pengukuran (*error score*).

Kesalahan pengukuran yang dimaksudkan dalam teori ini merupakan kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara teoritis dari skor amatan yang diperoleh dengan skor amatan yang diharapkan. Kesalahan pengukuran yang sistematis dianggap bukan merupakan kesalahan pengukuran.

Ada beberapa asumsi dalam teori tes klasik. Skor kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya, merupakan asumsi yang pertama. Asumsi yang kedua adalah skor kesalahan tidak berkorelasi dengan skor sebenarnya dan skor-skor kesalahan pada tes-tes yang lain untuk peserta tes (*testee*) yang sama. Ketiga, rata-rata dari skor kesalahan ini sama dengan nol. Asumsi-asumsi pada teori tes klasik ini dijadikan dasar untuk mengembangkan formula-formula dalam menentukan validitas dan reliabilitas tes.

Validitas dan reliabilitas pada perangkat tes digunakan untuk menentukan kualitas tes. Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran, daya pembeda dan efektivitas distraktor.

B. Reliabilitas

Mehrens & Lehmann (1973: 102) menyatakan bahwa reliabilitas merupakan derajat keajegan (*consistency*) di antara dua buah hasil pengukuran pada objek yang sama. Definisi ini dapat diilustrasikan dengan seseorang yang diukur tinggi badannya akan diperoleh hasil yang tidak berubah walaupun menggunakan alat pengukur yang berbeda dan skala yang berbeda. Dalam kaitannya dengan dunia pendidikan, prestasi atau kemampuan seorang siswa dikatakan reliabel jika dilakukan pengukuran, hasil pengukuran akan sama informasinya, walaupun penguji berbeda, korektornya berbeda atau butir soal yang berbeda tetapi memiliki karakteristik yang sama.

Allen & Yen (1979: 62) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya

dinyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Dalam pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri/karakter ini bersifat abstrak. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973: 103).

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Untuk melihat reliabilitas suatu alat ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan penelaahan secara statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

Untuk menentukan nilai reliabilitas suatu tes (butir soal berbentuk pilihan ganda (*multiple choice*)) dapat digunakan formula sebagai berikut

$$\hat{\alpha} = \frac{R}{R-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

dengan :

R : banyaknya butir soal,

σ^2 : varians.

Mehrens & Lehmann (1973: 104) menyatakan bahwa meskipun tidak ada perjanjian secara umum, tetapi secara luas dapat diterima bahwa untuk tes yang digunakan untuk membuat keputusan pada siswa secara perorangan harus memiliki koefisien reliabilitas minimal sebesar 0,85. Dengan demikian, pada penelitian ini, tes seleksi digunakan untuk menentukan keputusan pada siswa secara perorangan, sehingga indeks koefisien reliabilitasnya diharapkan minimal sebesar 0,85.

C. Validitas

Validitas suatu perangkat tes dapat diartikan merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur (Azwar, 2012: 51). Menurut Cohen-

Swerdlik (2009) *validity, as applied to a test, is a judgment or estimate of how well a test measures what it purports to measure in a particular context*. Pengertian ini menunjukkan bahwa validitas dalam sebuah tes menjadi hal yang sangat penting karena akan mengukur kemampuan peserta didik secara tepat.

Validitas juga merujuk pada ketepatan (*appropriateness*), kebermaknaan (*meaningfulness*) dan kemanfaatan (*usefulness*) kesimpulan yang didapatkan dari interpretasi skor tes (Kusaeri dan Suprananto, 2012). Ada tiga tipe validitas, yaitu validitas isi, validitas konstruk dan validitas kriteria (Cohen-Swerdlik 2009: 185; Azwar, 2012: 52).

Ada dua macam validitas isi, yaitu validitas kenampakan dan validitas logika (Azwar, 2012). Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan kemampuan yang hendak diukur (Azwar, 2012), yang berupa analisis rasional terhadap domain yang hendak diukur. Validitas kenampakan didasarkan pada pertanyaan apakah suatu butir-butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat.

Validitas kriteria, disebut juga validitas prediktif, merupakan kesahihan suatu perangkat tes dalam membuat prediksi, dapat meramalkan keberhasilan siswa pada masa yang akan datang. Validitas prediktif suatu perangkat tes dapat diketahui dari korelasi antara perangkat tes dengan kriteria tertentu yang dikehendaki, yang disebut dengan variabel kriteria (Cohen-Swerdlik, 2009: 97; Azwar, 2012).

D. Tingkat Kesukaran

Tingkat kesukaran suatu butir soal, yang disimbolkan dengan p_i , merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika p_i mendekati 0, maka soal tersebut terlalu sukar, sedangkan jika p_i mendekati 1, maka soal tersebut terlalu mudah, sehingga perlu

dibuang. Hal ini disebabkan karena butir tersebut tidak dapat membedakan kemampuan seorang siswa dengan siswa lainnya.

Sudjana (2011: 137) menyatakan bahwa secara umum indeks kesukaran suatu butir sebaiknya terletak dalam kategori sedang yakni pada interval 0,31 – 0,70. Pada interval ini, informasi tentang kemampuan siswa akan diperoleh secara maksimal. Dalam merancang indeks kesukaran suatu perangkat tes, perlu dipertimbangkan tujuan penyusunan perangkat tes tersebut. Untuk menentukan indeks kesukaran dari suatu butir pada perangkat tes pilihan ganda, digunakan persamaan sebagai berikut:

$$p_i = \frac{\sum B}{N}$$

dengan:

p = proporsi menjawab benar pada butir soal tertentu.

ΣB = banyaknya peserta tes yang menjawab benar.

N = jumlah peserta tes yang menjawab.

E. Daya Pembeda

Untuk menentukan daya pembeda, dapat digunakan indeks diskriminasi, indeks korelasi biserial, indeks korelasi *point biserial*, dan indeks keselarasan. Pada analisis butir dalam penelitian ini, hanya digunakan indeks korelasi *point biserial*. Koefisien korelasinya untuk suatu butir tes ditentukan dengan rumus:

$$r_{pbis} = \left[\frac{\bar{X}_1 - \bar{X}}{s_X} \right] \sqrt{\frac{p_1}{1-p_1}}$$

dengan r_{pbis} = koefisien korelasi point biserial, X_i merupakan variabel kontinyu, \bar{X}_1 merupakan rerata skor X untuk peserta tes yang menjawab benar butir tersebut, \bar{X} merupakan rerata skor X , s_X merupakan standar deviasi dari skor X , dan p_1 merupakan proporsi peserta tes yang menjawab benar butir tersebut.

Pada suatu butir soal, indeks daya beda dikatakan baik jika lebih besar atau sama dengan 0,3 (Nunnally & Bernstein, 2009: 304; Ebel & Frisbie, 1991; 232). Indeks daya pembeda suatu butir yang kecil nilainya akan menyebabkan butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Pada analisis tes dengan *Content-Referenced Measures*, indeks daya pembeda butir tidak terlalu perlu menjadi perhatian, asalkan tidak negatif (Ebel & Frisbie, 1991). Jika

nilainya kecil, menunjukkan bahwa kemencengan distribusi skor dari populasi, yang juga mengakibatkan validitas tes menjadi rendah.

F. Efektivitas Distraktor

Menurut Sudijono (2012) pada saat membicarakan tes objektif bentuk multiple choice *item* tersebut untuk setiap butir *item* yang dikeluarkan dalam tes hasil belajar telah dilengkapi dengan beberapa kemungkinan jawab, atau yang sering dikenal dengan istilah option atau alternatif.

Option atau alternatif itu jumlahnya berkisar antara 3 sampai dengan 5 buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir *item* itu, salah satu di antaranya adalah merupakan jawaban betul (kunci jawaban), sedangkan sisanya adalah merupakan jawaban salah. Jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (pengecoh).

Menganalisis fungsi distraktor sering dikenal dengan istilah lain, yaitu : menganalisis pola penyebaran jawaban *item*. Adapun yang dimaksud dengan pola penyebaran jawaban *item* adalah suatu pola yang dapat menggambarkan bagaimana *testee* menentukan pilihan jawabnya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir *item*.

Suatu kemungkinan dapat terjadi, yaitu bahwa dari keseluruhan alternatif yang dipasang pada butir *item* tertentu, sama sekali tidak dipilih oleh *testee*. Dengan kata lain, *testee* menyatakan “blangko”. Pernyataan blangko ini sering dikenal dengan istilah omiet dan biasa diberi lambang dengan huruf O.

Distraktor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5 % dari seluruh peserta tes.

Sebagai tindak lanjut atas hasil penganalisaan terhadap fungsi distraktor tersebut maka distraktor yang sudah dapat menjalankan fungsinya dengan baik dapat dipakai lagi pada tes-tes yang akan datang, sedangkan distraktor yang belum dapat berfungsi dengan baik sebaiknya diperbaiki atau diganti dengan distraktor yang lain.

DAFTAR PUSTAKA

- Azwar, Syaifuddin (2012) *Dasar-dasar Psikometri*. Yogyakarta: Pustaka Pelajar
- Daryanto (2005) *Evaluasi Pendidikan*. Jakarta: Rineka Cipta
- Depdiknas (2004). *Pedoman Umum Pengembangan Penilaian Kurikulum Berbasis Kompetensi SMA*. Jakarta: Direktorat Pendidikan Dasar dan Menengah, Direktorat Pendidikan Lanjutan Pertama.
- Depdiknas. (2006). *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*. Jakarta: Pusat Kurikulum, Balitbang Depdiknas
- Djaali & Mulyono, Pudji. (2007). *Pengukuran dalam Bidang Pendidikan*. Jakarta: Grasindo
- Ebel, Robert L. & David A. Frisbie (1991) *Essential Of Educational Measurement (5th Edition)*. New Delhi: Prentice-Hall, Inc.
- Gronlund, E. Norman. (1982). *Constructing Achievement Tests*. London: Prentice Hall.
- Hill, B.C., dan Ruptic, C.A. (1994). *Practical Aspects of Authentic Assessment: Putting the Pieces Together*. Norwood: Christopher-Gordon Publishers, Inc.
- Kusaeri dan Suprananto (2012) *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu
- Linn, R.L., dan Gronlund, N.E. (1995). *Measurement and Assessment in Teaching*. New Jersey: Prentice Hall.
- Nunnally, Jum C. & Ira H. Bernstein (1994) *Psychometric Theory (3rd Edition)*. New York: McGraw-Hill, Inc.
- Popham, W.J. (1995) *Classroom Assessment, What Teachers Need to Know*. Boston: Allyn & Bacon.
- Purwanti, Endang. (2008). *Asesmen Pembelajaran SD*. Direktorat Jendral Pendidikan Tinggi Departemen Pendidikan Nasional.
- Puskur (2008) *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*. Jakarta: Balitbang Depdiknas
- Slameto (2001) *Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Sudijono, Anas (2011) *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada
- Sudjana, Nana (2011) *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya
- Sukardi (2011) *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara
- Widoyoko, Eko Putro (2009) *Evaluasi Program Pembelajaran*. Diambil dari <http://www.umpwr.ac.id/web/download/publikasi-ilmiah> pada tanggal 22 Maret 2012