

KUALITAS INSTRUMEN TES HASIL BELAJAR

Khaerudin¹

khaerudin77@yahoo.com

Abstrak

Kualitas suatu tes hasil belajar sangat ditentukan oleh kualitas item-itemnya. Tes hasil belajar yang berisi item-item yang berkualitas tinggi walaupun dalam jumlah yang sedikit akan jauh lebih berguna dari pada tes hasil belajar yang berisi puluhan item berkualitas rendah yang akan menurunkan fungsi tes dan hasil pengukuran yang menyesatkan.

Pengolahan tes hasil belajar dalam rangka memperbaiki proses pembelajaran dapat di lakukan dengan membuat analisis soal (*item analysis*). Analisis kuantitatif dilakukan dengan cara mengujicobakan instrument yang telah dianalisis secara kualitatif kepada sejumlah siswa yang memiliki karakteristik sama dengan siswa yang akan diuji dengan instrument tersebut. Karakteristik internal secara kuantitatif dimaksudkan meliputi validitas, reliabilitas, daya pembeda, tingkat kesukaran serta efektifitas fungsi pengecoh (distraktor). Adapun manfaat atau kegunaan analisis soal buatan guru menurut Suke Silverius dalam bukunya yang berjudul Evaluasi Hasil Belajar dan Umpam Balik, adalah sebagai berikut: 1) Menentukan apakah butir soal berfungsi tepat seperti yang dimaksudkan oleh guru, 2) Umpam balik bagi siswa mengenai penampilannya dan merupakan dasar untuk diskusi kelas, 3) Umpam balik bagi guru tentang kesulitan belajar siswa, 4) Bidang-bidang kurikulum yang memerlukan perbaikan, 5) Perbaikan butir soal, dan 6) Meningkatkan ketrampilan penulisan soal.

Hasil belajar adalah tingkat penguasaan yang dicapai oleh siswa dalam mengikuti program pembelajaran sesuai dengan tujuan pendidikan yang telah ditetapkan.

Kata Kunci: Kualitas Instrumen Tes dan Hasil Belajar

A. Pendahuluan

Betapapun seringnya pergantian sistem evaluasi, namun kehadiran evaluasi dalam dunia pendidikan dan pengajaran tidak dapat dihindarkan. Keterjalinan hubungan antara tujuan pendidikan, kegiatan/proses pembelajaran, dan evaluasi

¹STIT Pemalang

adalah sedemikian erat sehingga tidak terpisahkan. Evaluasi diperlukan untuk memantau sejauh mana keberhasilan kegiatan pembelajaran dalam upayanya mencapai tujuan pendidikan.

Ada tiga alasan utama mengapa dalam kegiatan pendidikan selalu memerlukan evaluasi. *Pertama*, apabila dilihat dari pendekatan proses, adanya hubungan antara tujuan pendidikan, proses pembelajaran dan evaluasi. *Kedua*, kegiatan mengevaluasi terhadap hasil belajar merupakan salah satu ciri dari pendidik profesional. *Ketiga*, bila dilihat dari pendekatan kelembagaan, kegiatan pendidikan merupakan kegiatan manajemen, yang meliputi kegiatan *planning, programming, organizing, actuating, controlling, and evaluating*.

Dalam bidang pengajaran evaluasi bertujuan (1) menetapkan kompetensi isi pengajaran spesifik yang dimiliki oleh peserta didik, (2) memperbaiki proses belajar-mengajar. Sedangkan dalam bidang hasil belajar, evaluasi bertujuan (1) untuk mengetahui perbedaan kemampuan peserta didik, (2) untuk mengukur keberhasilan mereka baik secara individual maupun kelompok. Oleh karena itu, dibutuhkan instrumen-instrumen yang memadai. Sebab, keberhasilan mengungkapkan hasil dari pembelajaran siswa sebagaimana adanya (objektivitas hasil penilaian) sangat bergantung pada kualitas alat penilaiannya disamping pada cara pelaksanaannya.

Secara teoritis, siswa dalam suatu kelas merupakan populasi atau kelompok yang keadaannya heterogen, artinya setiap siswa mempunyai kemampuan yang berbeda dari siswa lainnya. Untuk itu, apabila dilakukan sebuah tes akan tercermin hasilnya dalam suatu kurva normal. Sebagian besar siswa berada di daerah sedang, sebagian kecil berada di ekor kiri, dan sebagian kecil yang lain berada di ekor kanan kurva. Namun, apabila keadaan setelah hasil tes dianalisis tidak seperti yang diharapkan dalam kurva normal, atau dengan kata lain apabila seluruh siswa (*testee*) memperoleh skor jelek atau sebaliknya apabila seluruh siswa memperoleh skor baik, maka tentu ada “sesuatu” dengan soal tesnya.

Kualitas suatu tes hasil belajar sangat ditentukan oleh kualitas item-itemnya. Tes hasil belajar yang berisi item-item yang berkualitas tinggi walaupun dalam jumlah yang sedikit akan jauh lebih berguna dari pada tes hasil belajar yang berisi puluhan item berkualitas rendah. Item-item yang berkualitas rendah tidak saja

menurunkan fungsi tes, tetapi juga akan memberikan hasil pengukuran yang menyesatkan.

Salah satu cara untuk menentukan kualitas suatu tes hasil belajar adalah dengan melakukan analisis soal (*item analysis*). Analisis soal terutama dapat dilakukan untuk tes objektif. Hal ini tidak berarti bahwa tes uraian tidak dapat dianalisis, akan tetapi memang dalam menganalisis butir tes uraian, belum ada pedoman yang standar. Jadi, tes hasil belajar bentuk objektif lebih mudah dianalisis dari pada tes hasil belajar bentuk uraian, baik dari segi validitas, reliabilitas, tingkat kesukaran maupun daya pembedanya.

Penilaian terhadap kegiatan dan hasil belajar mengajar siswa dimaksudkan untuk mengumpulkan data sebagai bahan pertimbangan dalam membantu perkembangan selanjutnya dan atau menetapkan keberhasilan siswa. Di samping penilaian itu, penilaian siswa merupakan bagian integral dari kegiatan pendidikan, dimaksudkan untuk memperoleh keterangan tentang kegiatan dan kemajuan belajar siswa.

B. Pembahasan

1. Kualitas Instrumen Tes

Salah satu cara untuk memperbaiki proses pembelajaran yang paling efektif ialah dengan jalan mengevaluasi tes hasil belajar yang di peroleh dari proses pembelajaran itu sendiri. Dengan kata lain, hasil tes itu kita olah sedemikian rupa sehingga dari hasil pengolahan itu dapat diketahui komponen-komponen manakah dari proses pembelajaran itu yang masih lemah.

Pengolahan tes hasil belajar dalam rangka memperbaiki proses pembelajaran dapat dilakukan dengan membuat analisis soal (*item analysis*).

1) Pengertian Analisis Item Tes

Analisis soal adalah suatu kegiatan yang berkaitan dengan proses mengumpulkan, meringkas, dan menggunakan informasi tentang jawaban siswa terhadap butir soal tes tersebut.² Nana Sudjana menyebutkan bahwa analisis item tes adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang

² Suke Silverius, *Evaluasi Hasil Belajar dan Umpam Balik*, (Jakarta: Grasindo, 1991), hlm. 166.

memadai.³ Menurut Saifuddin Azwar, analisis item tes adalah pengujian seluruh item tes yang didasarkan pada item empirik (data yang diperoleh dari hasil pengenaan tes yang sesungguhnya), agar diperoleh bukti mengenai kualitas item-item tes.⁴

Berdasarkan beberapa definisi di atas, maka dapat disimpulkan bahwa analisis item tes adalah proses pengkajian butir-butir tes hasil belajar yang didasarkan pada jawaban siswa terhadap tes tersebut, sehingga dapat diketahui kualitas dari suatu tes sebagai alat pengukur hasil belajar siswa.

2) Unsur-unsur Analisis Item Tes

Suatu instrumen hendaknya dianalisis sebelum digunakan. Ada dua model analisis yang dapat dilakukan, yaitu analisis kualitatif dan analisis kuantitatif. Analisis kualitatif adalah analisis yang dilakukan oleh teman sejawat dalam rumpun keahlian yang sama. Tujuannya adalah untuk menilai materi, kontruksi dan apakah bahasa yang digunakan sudah memenuhi pedoman dan sudah bisa dipahami oleh siswa.

Analisis kuantitatif dilakukan dengan cara mengujicobakan instrument yang telah dianalisis secara kualitatif kepada sejumlah siswa yang memiliki karakteristik sama dengan siswa yang akan diuji dengan instrument tersebut.⁵ Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris.⁶ Karakteristik internal secara kuantitatif dimaksudkan meliputi validitas, reliabilitas, daya pembeda, tingkat kesukaran serta efektifitas fungsi pengecoh (distraktor).

a. Validitas Tes

Validitas tes perlu ditentukan untuk mengetahui kualitas tes dalam kaitannya dengan mengukur hal yang seharusnya diukur. Kata “valid” diartikan dengan “tepat, benar, shahih, absah”. Jadi, kata

³ Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, (Bandung: Remaja Rosdakarya Offset, 1991), hlm. 135.

⁴ Saifuddin Azwar, *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*, (Yogyakarta: Pustaka Pelajar, 2000), hlm. 130

⁵ Abdul Majid, *Pendidikan Agama Islam Berbasis Kompetensi*, (Bandung: Remaja Rosdakarya, 2004), hlm. 223

⁶ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes*, (Bandung: Remaja Rosdakarya, 2005), hlm. 10.

validitas dapat diartikan dengan ketepatan, kebenaran, keshahihan atau keabsahan. Apabila kata valid itu dikaitkan dengan fungsi tes sebagai alat pengukur, maka sebuah tes dikatakan valid apabila tes tersebut dengan secara tepat, secara benar, secara shahih, atau secara absah dapat mengukur apa yang seharusnya diukur.⁷

Pengertian validitas menurut Sumarna Surapranata adalah “suatu konsep yang berkaitan dengan sejauhmana tes telah mengukur apa yang seharusnya diukur”.⁸ Menurut Mudjijo, suatu tes disebut valid apabila tes tersebut dapat mengukur apa yang hendak dan seharusnya diukur.⁹ Selanjutnya menurut Nana Sudjana, validitas adalah ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai.¹⁰

Dari beberapa definisi di atas, dapat disimpulkan bahwa suatu tes dapat dikatakan valid yaitu apabila tes tersebut dapat mengukur apa yang hendak dan seharusnya diukur.

Sedangkan yang dimaksud dengan validitas item tes adalah ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian yang tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.¹¹

a) Macam-macam Validitas

Validitas merupakan syarat yang terpenting dalam suatu alat evaluasi.¹² Untuk menentukan apakah suatu tes hasil belajar telah memiliki validitas atau daya ketepatan mengukur, dapat dilihat dari dua segi, yaitu dari segi tes itu sendiri sebagai totalitas (validitas tes), dan dari segi itemnya, sebagai bagian tak terpisahkan dari tes tersebut (validitas item tes).¹³

⁷ Anas Sudijono, *Pengantar Statistik Pendidikan*, (Jakarta: Rajawali, 1991), hlm. 93

⁸ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes, Op.Cit.*, hlm. 50.

⁹ Mudjijo, *Tes Hasil Belajar*, (Jakarta: Bumi Aksara, 1995), hlm. 40

¹⁰ Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar, Op.Cit.*, hlm. 12.

¹¹ Anas Sudijono, *Op.Cit.*, hlm. 182

¹² Ngalim Purwanto, *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*, (Bandung: Remaja Rosdakarya, 1994), hlm. 137.

¹³ Anas Sudijono, *Op.Cit.*, hlm. 163

(a) Validitas tes

Validitas sebuah tes dapat diketahui dari hasil pemikiran dan dari hasil pengalaman. Hal yang pertama akan diperoleh validitas logis (*logical validity*) dan hal yang kedua diperoleh validitas empiris (*empirical validity*). Dua hal inilah yang dijadikan dasar pengelompokan validitas tes adalah sebagai berikut: 1) Validitas logis. Validitas logis mengandung arti logis/penalaran, maka validitas logis untuk sebuah instrumen evaluasi menunjuk pada kondisi bagi sebuah instrumen yang memenuhi persyaratan valid berdasarkan hasil penalaran dan sudah dirancang secara baik, sesuai dengan teori dan ketentuan yang berlaku. Ada dua macam validitas logis yang dapat dicapai oleh sebuah instrumen, yaitu validitas isi dan validitas konstruksi, 2) Validitas Empiris. Dimaksud dengan validitas empiris adalah memiliki pengertian pengalaman, sehingga sebuah instrument dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Dengan demikian validitas empiris tidak dapat diperoleh hanya dengan jalan menyusun instrument berdasarkan ketentuan seperti halnya validitas logis, tetapi harus dibuktikan dengan hasil analisis yang dilakukan terhadap data hasil pengamatan dilapangan, terbukti bahwa tes hasil belajar itu dengan secara tepat telah dapat mengukur hasil belajar yang seharusnya diukur.¹⁴

Ada dua cara untuk mengetahui apakah tes hasil belajar itu sudah memiliki validitas empiris ataukah belum, yakni dari segi daya ketepatan meramalanya (*predictive validity*) dan daya ketepatan bandingannya atau “ada sekarang” (*concurrent validity*).¹⁵

¹⁴ Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2002), hlm. 65.

¹⁵ Anas Sudijono, *Op.Cit.*, hlm. 168

(b) Validitas item

Validitas item dari suatu tes adalah ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.¹⁶ Sebenarnya setiap butir item yang ada dalam tes hasil belajar itu adalah merupakan bagian tak terpisahkan dari tes hasil belajar tersebut sebagai suatu validitas dalam mengukur atau mengungkap hasil belajar yang telah dicapai oleh masing-masing individu peserta didik setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu.

b) Teknik Pengujian Validitas Item

Sebutir item dapat dikatakan telah memiliki validitas yang tinggi atau dapat dinyatakan valid, jika skor-skor pada butir soal yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya, atau dengan bahasa statistik: “Ada korelasi positif yang signifikan antara skor item dengan skor totalnya”. Skor total di siniberkedudukan sebagai variabel terikat (*dependent variable*), sedangkan skor item berkedudukan sebagai variabel bebasnya (*independent variable*).

Dengan demikian, maka untuk sampai pada kesimpulan bahwa butir-butir yang ingin diketahui validitasnya yaitu valid atau tidak kita dapat menggunakan teknik korelasi sebagai teknik analisisnya. Sebutir soal dapat dinyatakan valid, apabila skor butir yang bersangkutan terbukti mempunyai korelasi yang positif yang signifikan dengan skor totalnya. Seperti diketahui, pada tes objektif maka hanya ada dua kemungkinan jawaban, yaitu betul dan salah. Setiap butir soal yang dijawab dengan betul umumnya diberi skor 1 (satu), sedangkan untuk setiap jawaban yang salah diberikan skor 0 (nol). Jenis data seperti ini dalam dunia ilmu statistic dikenal dengan nama data diskret murni atau data dikotomik. Sedangkan skor total

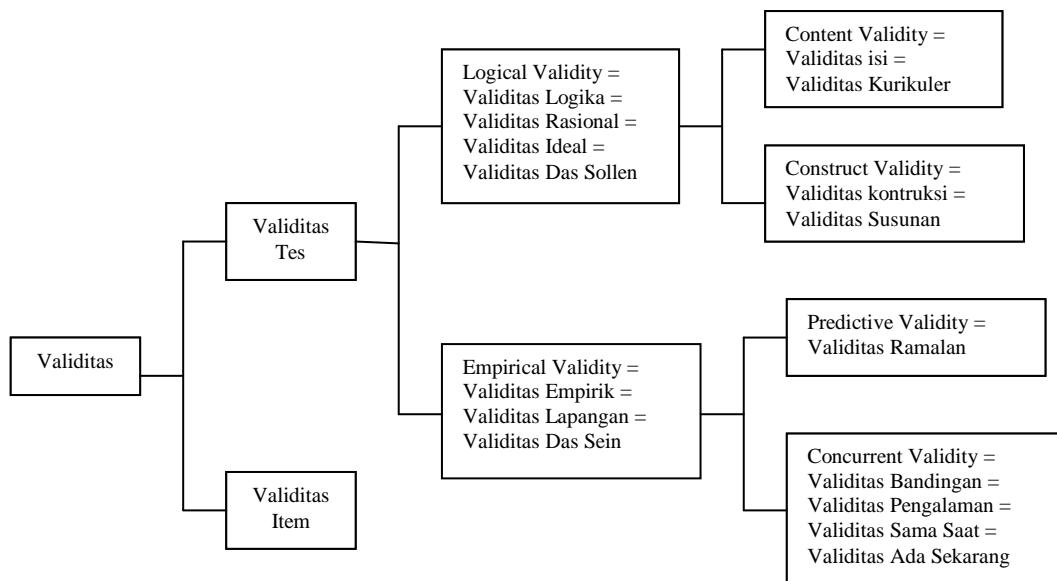
¹⁶ Anas Sudijono, *Op.Cit.*, hlm. 182.

yang dimiliki oleh masing-masing butir soal merupakan data kontinu.¹⁷

Sebuah item memiliki validitas yang tinggi, jika skor pada item mempunyai kesejajaran dengan skor total. Kesejajaran ini dapat diartikan dengan korelasi sehingga untuk mengetahui validitas item digunakan rumus korelasi.¹⁸

Menurut teori yang ada, apabila variabel I berupa data diskret murni atau data dikotomik (skor butir item), sedangkan variabel II berupa data kontinu (skor total butir item), maka teknik korelasi yang tepat untuk digunakan dalam mencari korelasi antara variabel I dengan variabel II adalah Teknik Korelasi Point Biserial, dimana indeks korelasinya diberi lambang (r_{pbi}).¹⁹

GAMBAR 2 BAGAN TENTANG VALIDITAS TES DAN VALIDITAS ITEM²⁰



¹⁷ *Ibid.*, hlm. 184-185

¹⁸ Suharsimi Arikunto, Dasar-Dasar Evaluasi Pendidikan, *Op.Cit.*, hlm. 162

¹⁹ Anas Sudijono, *Pengantar Statistik Pendidikan*, (Jakarta: Rajawali, 1991), hlm. 245

²⁰ Anas Sudijono, *Op. Cit.*, hlm. 185

(a) Reliabilitas Tes

Suatu tes yang reliabel memberikan suatu ukuran yang konsisten tentang kemampuan siswa untuk mempertanyakan prestasi mengenai suatu tujuan. Reliabilitas menunjukkan nilai-nilai yang konsisten. Suatu instrumen yang mempunyai tingkat reliabilitas yang tinggi dapat dipercaya untuk dijadikan dasar pengambilan kesimpulan dan keputusan.

Reliabilitas merupakan penerjemahan dari kata reliability yang mempunyai asal kata *rely* dan *ability*. Reliabilitas sering diartikan dengan keterandalan. Artinya suatu tes memiliki keterandalan bilamana tes tersebut dipakai mengukur berulang-ulang hasilnya sama.²¹ Reliabilitas adalah ketatapan atau ketelitian suatu alat evaluasi. Suatu tes atau alat evaluasi dikatakan andal jika ia dapat dipercaya, konsisten, atau stabil dan produktif.²² Jadi, yang dipentingkan di sini adalah ketelitiannya, sejauhmana tes atau alat tersebut dapat dipercaya kebenarannya.

Tes Hasil belajar dikatakan baik apabila telah memiliki reliabilitas atau bersifat reliabel. Apabila istilah tersebut dikaitkan dengan fungsi tes sebagai alat ukur mengenai keberhasilan belajar peserta didik, maka sebuah tes tersebut dapat dinyatakan reliable apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama, senantiasa menunjukkan hasil yang tetap sama atau sifatnya ajeg dan stabil.²³ Ajeg atau tetap di sini tidak selalu harus sama, tetapi mengikuti perubahan secara ajeg.

Dari beberapa definisi di atas, maka hasil pengukuran dapat dipercaya hanya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama diperoleh

²¹ M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, (Jakarta: Raja Grafindo Persada, 1996), hlm. 118.

²² Ngalim Purwanto, *Op.Cit.*, hlm.139

²³ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, *Op.Cit.*, hlm. 95

hasil relatif sama, selama aspek yang diukur dalam diri subyek yang diukur memang belum berubah.

1) Faktor-faktor yang Mempengaruhi Reliabilitas

Beberapa faktor yang mempengaruhi reliabilitas adalah sebagai berikut: a) Luas tidaknya sampling yang handal, b) Perbedaan bakat dan kemampuan murid yang dites, dan c) Suasana dan kondisi testing.²⁴

2) Teknik Pengujian Reliabilitas

Dalam rangka menentukan apakah tes hasil belajar bentuk obyektif yang disusun oleh tester telah memiliki keajegan mengukur ataukah belum, dapat dilakukan dengan menggunakan tiga teknik yang dapat digunakan untuk menguji tingkat reliabilitas butir tes, yaitu:

a) Metode *Test-retest* (metode bentuk ulang)

Metode ini digunakan untuk menguji dengan menggunakan alat penilaian terhadap subyek yang sama, dilakukan dua kali dalam waktu yang berlainan, kemudian dikorelasikan.²⁵ Koefisien korelasi yang diperoleh menunjukkan tingkat konsistensi instrumen yang sekaligus juga merupakan nilai koefisien korelasi. Hasil uji teknik ini dapat dipercaya bila instrumen tersebut mengukur variabel yang relatif konstan.

Adapun langkah yang dapat ditempuh pada uji reliabilitas ini adalah sebagai berikut: 1) Menyusun sebuah tes yang akan diukur reliabilitasnya, 2) Mengujikan tes yang tersusun tersebut (tahap I), 3) Menghitung skor hasil tes tahap I, 4) Mengujikan ulang tes yang tersusun tersebut (tahap II), 5) Menghitung skor hasil tes ulang (tahap II), dan 6) Menghitung reliabilitas tes tersebut dengan jalan mengorelasikan skor tes I

²⁴Ngalim Purwanto, *Op.Cit.*, hlm. 141

²⁵Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, *Op.Cit.*, hlm. 17.

dengan skor tes II dengan rumus *Korelasi Product Moment Person*.²⁶

b) Pendekatan *equivalent-form reliability* (bentuk paralel)

Pendekatan ini dipakai dengan dua bentuk tes yang sama yang dilaksanakan oleh satu kelompok pada waktu yang sama. Bentuk-bentuk tes itu sama dalam arti, bahwa tes itu disusun untuk mengukur kemampuan yang sama.²⁷ Cara ini dapat digunakan untuk mengetahui koefisien stabilitas tes dengan asumsi bahwa sistem yang diukur dengan tes tersebut tidak akan berubah dengan hanya digunakan dengan dua bentuk tes. Adapun langkah yang ditempuh adalah sebagai berikut: 1) Menyusun dua buah tes yang ekuivalen, 2) Menyusun kedua tes tersebut (dalam waktu yang bersamaan atau beriringan), 3) Memberikan skor hasil tes yang telah diujikan, disusun dengan memisahkan antara tes A dengan tes B, dan 4) Mencari koefisien stabilitas kedua tes (A dan B) dengan jalan mencari korelasinya melalui rumus *Korelasi Product Moment*.²⁸

c) Pendekatan *split-half*

Metode ini dipakai untuk digunakan dalam rangka menentukan reliabilitas dengan jalan melakukan pengukuran terhadap satu kelompok subyek, dimana pengukuran itu dilakukan dengan hanya menggunakan satu jenis alat pengukur, sedang pelaksanaan pengukuran itu hanya dilakukan sebanyak satu kali saja. Dengan kata lain metode ini dilakukan dengan satu kelompok subyek, satu jenis alat pengukur dan satu kali pengukuran; atau satu kelompok testee, satu jenis tes, dan satu kali

²⁶Chabib Thoha, *Op.Cit.*, hlm. 120.

²⁷ Norman F. Gronlund, *Constructing Achievement Test*, (Semarang: IKIP Press, 1981), hlm. 242.

²⁸ *Ibid*, hlm. 123

testing.²⁹ Adapun langkah secara umum yang ditempuh untuk mencari reliabilitas tes adalah: 1) Menyusun sebuah tes setidaknya jumlah nomornya genap, sehingga bila dibelah jumlahnya sama, 2) Mengujikan tes tersebut pada satu sampel, 3) Menghitung skor masing-masing peserta didik dalam dua kelompok skor, dapat dikelompokkan skor ganjil dan genap, dapat pula dikelompokkan skor belahan atas dan skor belahan bawah, 4) Mencari reliabilitas setengah tes dengan jalan mengkorelasikan kedua skor tersebut dengan rumus *Product Moment* atau mencari deviasi pada belahan ganjil genap, dan 5) Mencari reliabilitas satu tes penuh dengan menggunakan rumus *Spearman Brown*, rumus *Flanagan* dan rumus *Rulon*.³⁰

d) Internal konsistensi (*internal consistency*)

Untuk mengukur koefisien konsistensi dapat digunakan pendekatan yang tidak membelah tes menjadi dua. Hal ini disebabkan oleh dua kemungkinan; 1) jumlah item ganjil, sehingga tidak dapat dibelah menjadi dua, 2) komposisi antara item-item ganjil dan genap tidak homogen, sehingga bila dibelah cenderung tidak memiliki korelasi yang positif.³¹

Internal konsisten yang didasarkan pada homogenitas atau korelasi antar skor jawaban pada setiap butir tes. Jika korelasi rerata antar butir soal tinggi maka reliabilitasnya juga tinggi. Jika korelasi rerata mendekati nol. Maka internal konsistensi nol pula dan reliabilitasnya rendah. Terdapat beberapa teknik dan persamaan yang digunakan untuk mencari reliabilitas dengan internal konsistensi ini yaitu; 1) koefisien *alpha*,

²⁹ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, Op.Cit., hlm. 214

³⁰ Chabib Thoha, *Op.Cit.*, hlm. 124.

³¹ *Ibid*, hlm. 133.

2) Kuder-Richardson₂₀, 3) Kuder-Richardson₂₁, dan 4) teknik Hoyt.³²

TABEL I
METODE UNTUK MENENTUKAN RELIABILITAS³³

Bentuk Reliabel	Prosedur untuk Memperoleh
<i>Test – retest methods</i> (stabilitas): <i>Product Moment</i> dan Korelasi Intra Kelas	Sajikan tes yang sama sebanyak dua kali kepada peserta tes yang sama dalam waktu berbeda dan tentukan korelasi
Paralel Ekuivalen: Produk momen dan korelasi intra kelas	Sajikan dua tes yang sama kepada peserta tes yang sama dalam waktu yang relatif tidak lama (misalnya dua minggu), korelasikan kedua skor tersebut untuk mencari Reliabilitas
<i>Split-Half methods</i> (belah dua) Persamaan <i>Split-Half</i> dan <i>Spearman – Brown</i>	Sajikan satu kali tes lalu di belah dua, gunakan persamaan untuk mengkorelasikan kedua belahan
<i>Internal Consistency</i> <input type="checkbox"/> Koefisien <i>alpha</i> <input type="checkbox"/> Kuder-Richardson (KR-20) <input type="checkbox"/> Kuder-Richardson (KR-21)	Berikan sekali tes, gunakan persamaan Berikan sekali tes, gunakan persamaan Berikan sekali tes, gunakan persamaan

Dari uraian yang telah dikemukakan diatas, kiranya menjadi cukup jelas. Kemudian langkah pengujian reliabilitas yang akan dilakukan dalam penelitian ini adalah penulis akan menggunakan rumus *koefisien alpha*.

(b) Tingkat Kesukaran Item

1) Pengertian Tingkat Kesukaran Item

Tingkat kesukaran Item adalah pernyataan tentang seberapa mudah dan seberapa sulit sebuah butir soal bagi

³² Sumarna Surapranata, *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes*, *Op.Cit.*, hlm. 113.

³³ *Ibid.*, hlm. 91

siswa yang dikenai pengukuran.³⁴ Suke Silverius menyebutkan bahwa tingkat kesukaran item adalah persentase siswa yang dapat menjawab benar butir soal tersebut.³⁵

Dari beberapa pengertian di atas, dapat disimpulkan tingkat kesukaran adalah seberapa besar tingkat kesulitan/kesukaran suatu butir soal yang ditunjukkan dengan persentase siswa yang menjawab benar terhadap butir soal tersebut.

2) Teknik Analisis Tingkat Kesukaran

Asumsi yang digunakan untuk memperoleh kualitas soal yang baik, disamping memenuhi validitas dan reliabilitas adalah adanya keseimbangan dari tingkat kesulitan soal tersebut. Keseimbangan yang dimaksudkan adalah adanya soal-soal yang termasuk mudah, sedang, dan sukar secara proporsional. Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dilihat dari sudut guru sebagai pembuat soal.³⁶

Secara *tentatif* dapat dikatakan bahwa salah satu ciri butir soal yang baik adalah bahwa ia tidak terlalu sukar dan tidak terlalu mudah untuk kelompok tertentu yang akan dites.³⁷

Cara yang dapat ditempuh untuk mengetahui apakah item tes hasil belajar itu sudah memiliki tingkat kesukaran yang memadai ataukah belum, maka dapat diketahui dari besar kecilnya indeks kesukaran item (*difficulty index*).

Indeks kesukaran item adalah bilangan atau angka yang menunjukkan sukar dan mudahnya suatu item soal.

³⁴ Burhan Nurgiyanto, *Penilaian Dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta: BPFE, 1987), hlm. 126

³⁵ Suke Silverius, *Op.Cit.*, hlm. 167.

³⁶ Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, *Op.Cit.*, hlm. 135

³⁷ Mudjijo, *Op.Cit.*, hlm. 62.

Besarnya indeks kesukaran adalah antara 0,00 sampai dengan 1,00. Artinya suatu soal yang indeks kesukarannya 0,00 menunjukkan bahwa soal itu terlalu sukar, sebaliknya suatu soal yang indeks kesukarannya 1,00 menunjukkan bahwa soal tersebut terlalu mudah. Dalam istilah evaluasi, indeks kesukaran diberi simbol "p" singkatan dari proporsi. Untuk mencari bilangan indeks kesukaran (p), digunakan rumus sebagai berikut: a) Soal yang memiliki $p < 0,30$ adalah soal yang sukar, b) Soal yang memiliki $0,30 \leq p \leq 0,70$ adalah soal yang sedang, dan c) Soal yang memiliki $p > 0,70$ adalah soal yang mudah.³⁸

(c) Daya Pembeda

1) Pengertian Daya Pembeda Item

Daya pembeda suatu soal tes adalah bagaimana kemampuan soal itu untuk membedakan siswa-siswa yang termasuk kelompok pandai (*upper group*) dengan siswa-siswa yang termasuk kelompok kurang (*lower group*).³⁹

Sedangkan Anas Sudijono menjelaskan bahwa, daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara testee yang berkemampuan tinggi (pandai) dengan testee yang kemampuannya rendah (bodoh) demikian rupa, sehingga sebagian besar testee yang memiliki kemampuan tinggi untuk menjawab butir item tersebut lebih banyak yang menjawab betul, sementara *testee* yang kemampuannya rendah untuk menjawab butir item tersebut, sebagian besar tidak dapat menjawab item dengan betul.⁴⁰

Dari beberapa definisi di atas, maka dapat disimpulkan bahwa suatu item tes memiliki daya pembeda yaitu apabila item tes itu dapat dijawab benar oleh siswa kelompok atas

³⁸ Sumarna Surapranata, *Op.Cit.*, hlm. 19.

³⁹ Ngalim Purwanto, *Op.Cit.*, hlm. 120.

⁴⁰ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, *Op.Cit.*, hlm. 385-386.

(pandai) dan tidak dapat dijawab benar oleh siswa kelompok bawah (bodoh).

2) Teknik Analisis Daya Pembeda

Indeks daya pembeda dihitung atas dasar pembagian kelompok menjadi dua bagian, yaitu kelompok atas yang merupakan peserta tes yang berkemampuan tinggi dengan kelompok bawah, yaitu kelompok peserta tes yang berkemampuan rendah. Indeks daya pembeda didefinisikan sebagai selisih antara proporsi jawaban benar pada kelompok atas dengan proporsi jawaban benar pada kelompok bawah. Pembagian kelompok ini dapat dilakukan dengan metode yang paling banyak dipakai adalah dengan menentukan 27% kelompok atas dan 27% kelompok bawah.⁴¹

Daya pembeda item dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi. Indeks diskriminasi item pada umumnya diberi lambang "D" (*discriminatory power*). Sebagaimana indeks kesukaran, indeks diskriminasi ini berkisar antara 0,00 sampai dengan 1,00. Dalam indeks diskriminasi tanda negatif digunakan jika suatu soal terbalik menunjukkan kualitas testee, yaitu anak pandai disebut bodoh dan anak bodoh disebut pandai.

Dengan demikian, interpretasi indeks daya beda yang digunakan adalah sebagai berikut:

D : 0,00 – 0,20 = jelek

D : 0,20 – 0,40 = cukup

D : 0,40 – 0,70 = baik

D : 0,70 – 1,00 = baik sekali

D : negatif (-) = tidak baik.⁴²

⁴¹ Anas Sudijono, *Op. Cit.*, hlm. 387.

⁴² Suharsimi Arikunto, *Op. Cit.*, hlm. 218.

(d) Efektifitas Fungsi Distraktor

Dalam setiap tes objektif selalu digunakan alternatif jawaban yang mengandung dua unsur sekaligus, yaitu jawaban tepat dan jawaban yang salah sebagai penyesat (*distraktor*).⁴³

Tujuan utama pemasangan distraktor pada setiap butir item itu adalah, agar dari sekian banyak testee yang mengikuti tes hasil belajar ada yang tertarik atau terangsang untuk memilihnya, sebab mereka menyangka bahwa distraktor yang mereka pilih itu merupakan jawaban betul. Makin banyak testee yang terkecoh, maka distraktor tersebut dapat menjalankan fungsinya dengan baik. Sebaliknya, apabila distraktor yang dipasang pada setiap butir item itu "tidak laku" (maksudnya: tidak ada seorangpun dari sekian banyak testee yang merasa tertarik atau terangsang untuk memilih distraktor tersebut sebagai jawaban betul), maka distraktor tersebut tidak dapat menjalankan fungsinya dengan baik. Dengan kata lain, distraktor baru dapat dikatakan telah menjalankan fungsinya dengan baik, apabila distraktor tersebut memiliki daya rangsang atau daya tarik, sehingga testee (khususnya testee dari kelompok bawah) menjadi terkecoh untuk memilih distraktor sebagai jawaban betul.⁴⁴

Analisis fungsi distraktor yang sering dikenal dengan istilah lain, yaitu pola penyebaran jawaban soal. Adapun yang dimaksud pola penyebaran jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih option a, b, c, atau d atau yang tidak memilih option manapun (blangko). Dalam istilah evaluasi disebut omit, disingkat O.⁴⁵

⁴³ Chabib Thoha, *Op.Cit.*, hlm. 149.

⁴⁴ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, *Op.Cit.*, hlm. 410.

⁴⁵ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, *Op.Cit.*, hlm. 219.

Suatu distraktor dapat diperlakukan dengan 3 cara: 1) Diterima, karena sudah baik, 2) Ditolak, karena tidak baik, dan 3) Ditulis kembali, karena kurang baik.⁴⁶

3) Kegunaan Analisis Item Tes

Analisis item tes (*item analysis*) merupakan suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun. Suharsimi Arikunto dalam bukunya yang berjudul Dasar-Dasar Evaluasi Pendidikan, menjelaskan bahwa faedah atau kegunaan dari analisis item tes adalah: a) Membantu kita dalam mengidentifikasi butir-butir soal yang jelek, b) Memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan lebih lanjut, dan c) Memperoleh gambaran secara selintas tentang keadaan tes yang kita susun.⁴⁷

Adapun manfaat atau kegunaan analisis soal buatan guru menurut Suke Silverius dalam bukunya yang berjudul Evaluasi Hasil Belajar dan Umpan Balik, adalah sebagai berikut:⁴⁸

- a. Menentukan apakah butir soal berfungsi tepat seperti yang dimaksudkan oleh guru.

Untuk menentukan apakah butir soal telah berfungsi sebagaimana mestinya, guru perlu mempertimbangkan hal-hal berikut: 1) Apakah tes itu ditujukan untuk mengukur pencapaian tujuan instruksional yang dimaksudkan?, 2) Apakah tes itu mempunyai tingkat kesukaran yang memadai, dipandang dari materi yang dipakai untuk menulis butir soal itu dan tingkat kemampuan yang diukur?, 3) Apakah kunci jawaban telah betul?, dan 4) Apakah distraktor berfungsi dengan baik?

- b. Umpan balik bagi siswa mengenai penampilannya dan merupakan dasar untuk diskusi kelas.

Siswa berhak mengetahui bagaimana tesnya dinilai dan jawaban yang benar dari setiap butir soal. Dengan demikian dia dapat membetulkan kesalahan jawabannya, sementara guru dapat menjelaskan

⁴⁶ Daryanto, *Op.Cit.*, hlm. 193

⁴⁷ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, *Op.Cit.*, hlm. 205.

⁴⁸ Suke Silverius, *Op.Cit.*, hlm. 176-177.

sejauhmana jawaban yang diinginkan dari setiap soal. Hal ini menyebabkan siswa lebih memahami pokok bahasan atau subpokok bahasan melalui jawaban yang baik dan benar dari setiap soal.

- c. Umpam balik bagi guru tentang kesulitan belajar siswa.

Suatu prosedur sederhana seperti mentabulasi presentase siswa yang menjawab benar suatu butir soal dapat memberikan informasi kepada guru mengenai pokok-pokok bahasan yang membutuhkan penjelasan tambahan dan perbaikan. Tentu saja sekelompok butir soal yang menanyakan bahan yang sama akan memberikan informasi yang lebih reliabel (ajeg) daripada satu soal saja.

Mengidentifikasi kesalahan apa yang ada dalam jawaban terhadap soal-soal dapat sangat membantu guru untuk perbaikan tingkat pemahaman siswa terhadap pokok bahasan atau subpokok bahasan yang diteskan itu.

- d. Bidang-bidang kurikulum yang memerlukan perbaikan.

Jika ada butir soal tertentu yang selalu sukar bagi siswa, atau selalu ada jenis kesalahan tertentu yang sering terjadi, maka mungkin masalahnya di luar jangkauan guru-guru. Mungkin kurikulumnya yang perlu direvisi. Analisis soal dapat membantu menemukan hal ini.

- e. Perbaikan butir soal.

Hasil analisis butir soal dapat menunjukkan kualitas butir soal itu. Maka hasil analisis dapat dipakai untuk mengupayakan perbaikan butir soal tersebut. Butir-butir soal yang diperbaiki itu dapat disimpan untuk dipakai lagi pada tahun yang akan datang.

- f. Meningkatkan ketrampilan penulisan soal.

Cara yang paling efektif untuk meningkatkan keterampilan menulis soal tes adalah menganalisis butir-butir soal dan cara siswa menjawab soal-soal itu. Kemudian, memanfaatkan informasi ini untuk perbaikan butir soal dan mencobanya lagi kepada para siswa. Hanya membaca buku teori tidaklah cukup.

2. Hasil Belajar

Kemampuan siswa untuk menampilkan berbagai aktivitas yang diharapkan, dimana kegiatan tersebut harus mereka pelajari melalui kegiatan instruksional disebut belajar. Hasil belajar adalah tingkat penguasaan yang dicapai oleh siswa dalam mengikuti program pembelajaran sesuai dengan tujuan pendidikan yang telah ditetapkan.⁴⁹ Dalam *Taxonomy of Educational Objectives*, Bloom mengklasifikasikan hasil belajar menjadi tiga ranah, yaitu: 1) Ranah kognitif, 2) Ranah afektif, dan 3) Ranah psikomotor.⁵⁰

Briggs mengemukakan bahwa hasil belajar adalah seluruh kecakapan dan hasil yang dicapai melalui proses belajar. Mengajar di sekolah yang dinyatakan dengan angka atau nilai yang diukur dengan tes hasil belajar.⁵¹ Demikian juga Gronlund mengemukakan, hasil belajar adalah sebuah prosedur sistematis untuk menentukan berapa banyak yang telah dipelajari seorang siswa.⁵² Lebih lanjut Nitko mengatakan bahwa, hasil belajar adalah prosedur sistematis untuk mendapatkan informasi yang digunakan dalam pengambilan keputusan tentang pelajar, kurikulum dan program serta kebijakan pendidikan dengan mengamati dan mendeskripsikan satu, atau lebih karakteristik menggunakan skala numerik atau skema klasifikasi.

Hasil belajar adalah kemampuan-kemampuan yang dimiliki siswa setelah ia menerima pengalaman belajarnya. Kemampuan-kemampuan yang dimiliki siswa akibat perbuatan belajar dan dapat diamati melalui penampilan siswa (*learner's performance*).⁵³ Penampilan siswa yang dimaksudkan tersebut adalah kemampuan yang dicapai dan diaplikasikan oleh siswa dalam merespon setiap obyek yang dihadapi. Dick dan Reiser menjelaskan bahwa hasil belajar

⁴⁹ Robert M.Gagne and Leslie J.Briggs,*Principles of Instructional Design*, (New York: Holt Rinehart and Winston, 1974), hlm. 47.

⁵⁰ Benjamin S. Bloom, J. Thomas Hastings and George F. Madaus,*Handbook on Formative and Summative Evaluation of Student Learning*, (New York: McGraw-Hill Book Co.,1971), hlm.271-277.

⁵¹ Leslie J. Briggs,*Instruksional Design Principles and Applications*, (New Jersey: Englewood Cliffs, 1993), hlm. 149.

⁵² Norman Gronlund, *Constructing Achievement Test*, (Englewood Cliffs: Prentice Hall, 1982), hlm. 1.

⁵³ Robert M. Gagne and Marcy P. Driscoll,*Essential of Learning for Instruction*, (New York: Prentice Hall, Inc., 1988), hlm.36.

adalah kemampuan-kemampuan yang dimiliki siswa sebagai hasil kegiatan pembelajaran.⁵⁴

Kingsley membagi hasil belajar atas tiga macam, yakni: 1) Keterampilan dan kebiasaan, 2) Pengetahuan dan pengertian, dan 3) Sikap dan ciri-ciri. Hasil belajar itu diperoleh siswa setelah mengikuti kegiatan belajar. Untuk mengetahui tingkat pencapaian hasil belajar siswa, guru menggunakan tes hasil belajar dan biasanya dinyatakan dalam bentuk skor.⁵⁵

Gagne *dalam* Burhan mengemukakan bahwa kompetensi dan kapabilitas sebagai bukti nyata hasil belajar dan dapat dibedakan kedalam lima kategori: 1) Keterampilan Intelektual (*Intellectual Skill*), 2) Strategi Kognitif (*Cognitif Strategy*), 3) Informasi Verbal (*Verbal Information*), 4) Keterampilan Motorik (*Motor Skill*), dan 5) Sikap (*Attitude*).⁵⁶

Selama siswa belajar, ia akan dihadapkan pada soal-soal untuk dipecahkan dan diatasi (*Problem Solving*). Suatu masalah dapat diartikan sebagai soal yang harus diselesaikan. Pemecahan masalah merupakan sesuatu yang terpadu dalam diri pembelajar dan hasil belajar.

Menurut Polya (*dalam* Setiabudi) ada empat langkah yang perlu dilakukan dalam menyelesaikan suatu soal, yaitu: 1) Memahami soal yang ada, 2) Menyusun suatu strategi, 3) Melakukan strategi yang telah dipilih dan 4) Menilai kembali pekerjaan yang telah dilakukan.⁵⁷

Untuk dapat membedakan antara Tes Hasil Belajar (THB) dengan Tes Prestasi Belajar (TPB), maka Nasoetion dan Suryanto menjelaskan sebagai berikut: Tes Hasil Belajar (THB) adalah alat ukur yang mampu menentukan kemampuan seseorang setelah mengikuti pembelajaran. Materi yang dinyatakan tidak hanya mengenai materi yang diperoleh dari guru saja tetapi juga mengenai hal-hal diluar yang diberikan, dilatihkan dan didiskusikan dengan guru, sedangkan Tes Prestasi Belajar (TPB) adalah alat ukur yang mampu

⁵⁴ Walter Dick and Robert A. Reiser, *Planing Effective Instruction*, (Boston: Allyn and Bacon, 1988). hlm.11

⁵⁵ Gorry H. Kingsey, *The Nature and Conditioning of Learning*, (New Jersey: Englewood Cliffs, 1970), hlm. 15.

⁵⁶ Burhan Nurgianto, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta: BPFE, 1995), hlm. 22-24.

⁵⁷ Wono Setya Budi, *Langkah Awal Menuju Olimpiade Matematika*, (Jakarta: Ricardo, 2003), hlm. 2.

menentukan seberapa banyak pelajaran yang telah diikuti dapat dikuasai/diserap oleh peserta didik.⁵⁸

Perlu ditekankan bahwa penilaian hasil belajar siswa tidak hanya menyangkut aspek-aspek kognitifnya saja, tetapi juga mengenai aplikasi atau *performance*, aspek afektif yang menyangkut sikap internalisasi nilai-nilai yang perlu ditanamkan dan dibina melalui materi yang telah diberikan.⁵⁹

Dari beberapa teori mengenai pengertian tentang hasil belajar diatas, maka dapat disimpulkan bahwa hasil belajar adalah prosedur sistematis untuk mendapatkan informasi yang digunakan dalam pengambilan keputusan yang dinyatakan dalam nilai atau angka berdasarkan hasil yang dicapai melalui proses belajar. Hasil belajar yang dicapai siswa melalui proses pembelajaran sebagai berikut: 1) Menambah keyakinan atas kemampuan dirinya dalam belajar, 2) Termotivasinya pribadi siswa secara intrinsik, 3) Menyadari bahwa hasil belajar yang dicapai sangat bermakna bagi dirinya, 4) Kemampuannya untuk dapat mengontrol atau menilai dan mengendalikan dirinya terutama dalam menilai hasil yang telah diperolehnya, dan 5) Hasil belajar diperoleh secara menyeluruh (*komprehensif*).

C. Kesimpulan

Salah satu cara untuk memperbaiki proses pembelajaran yang paling efektif ialah dengan jalan mengevaluasi tes hasil belajar yang di peroleh dari proses pembelajaran itu sendiri. Dengan kata lain, hasil tes itu kita olah sedemikian rupa sehingga dari hasil pengolahan itu dapat di ketahui komponen-komponen manakah dari proses pembelajaran itu yang masih lemah.

Tes Hasil belajar dikatakan baik apabila telah memiliki reliabilitas atau bersifat reliabel. Apabila istilah tersebut dikaitkan dengan fungsi tes sebagai alat ukur mengenai keberhasilan belajar peserta didik, maka sebuah tes tersebut dapat dinyatakan reliable apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama, senantiasa menunjukkan hasil yang tetap sama atau sifatnya ajeg dan stabil.

⁵⁸ Noehi Nasution dan Adi Suryanto, *Evaluasi Pengajaran*, (Jakarta: Universitas Terbuka, 2000), hlm. 1.4.

⁵⁹ Oemar Hamalik, *Perencanaan Pengajaran Berdasarkan Pendekatan Sistem*, (Jakarta: Bumi Aksara, 2003), hlm. 211.

Hasil belajar adalah prosedur sistematis untuk mendapatkan informasi yang digunakan dalam pengambilan keputusan yang dinyatakan dalam nilai atau angka berdasarkan hasil yang dicapai melalui proses belajar. Hasil belajar yang dicapai siswa melalui proses pembelajaran sebagai berikut: 1) Menambah keyakinan atas kemampuan dirinya dalam belajar, 2) Termotivasi pribadi siswa secara intrinsik, 3) Menyadari bahwa hasil belajar yang dicapai sangat bermakna bagi dirinya, 4) Kemampuannya untuk dapat mengontrol atau menilai dan mengendalikan dirinya terutama dalam menilai hasil yang telah diperolehnya, dan 5) Hasil belajar diperoleh secara menyeluruh (*komprehensif*).

Daftar Pustaka

- Arikunto, Suharsimi. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2002.
- Azwar, Saifuddin. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar, 2000.
- Bloom, Benjamin S., J. Thomas Hastings and George F. Madaus. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill Book Co., 1971.
- Briggs, Leslie J. *Instruksional Design Principles and Applications*. New Jersey: Englewood Cliffs, 1993.
- Budi, Wono Setya. *Langkah Awal Menuju Olimpiade Matematika*. Jakarta: Ricardo, 2003.
- Dick, Walter and Robert A. Reiser. *Planing Effective Instruction*. Boston: Allyn and Bacon, 1988.
- Gagne, Robert M. and Leslie J. Briggs. *Principles of Instructional Design*. New York: Holt Rinhart and Winston, 1974.
- Gagne, Robert M. and Marcy P. Driscoll. *Essensial of Learning for Instruction*. New York: Prentice Hall, Inc., 1988.

- Gorry H. Kingsey. *The Nature and Conditioning of Learning*. (New Jersey: Englewood Cliffs, 1970), hlm. 15.
- Gronlund, Norman. *Constructing Achievement Test*. Englewood Cliffs: Prentice Hall, 1982.
- Hamalik, Oemar. *Perencanaan Pengajaran Berdasarkan Pendekatan Sistem*. Jakarta: Bumi Aksara, 2003.
- Majid, Abdul. *Pendidikan Agama Islam Berbasis Kompetensi*. Bandung: Remaja Rosdakarya, 2004.
- Nasution, Noehi dan Adi Suryanto. *Evaluasi Pengajaran*. Jakarta: Universitas Terbuka, 2000.
- Nurgianto, Burhan. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE, 1995.
- Nurgiyanto, Burhan. *Penilaian Dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE, 1987.
- Purwanto, Ngahim. *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya, 1994.
- Silverius, Suke. *Evaluasi Hasil Belajar dan Umpulan Balik*. Jakarta: Grasindo, 1991.
- Sudijono, Anas. *Pengantar Statistik Pendidikan*. Jakarta: Rajawali, 1991.
- Sudijono, Anas. *Pengantar Statistik Pendidikan*. Jakarta: Rajawali, 1991.
- Sudjana, Nana. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya Offset, 1991.
- Surapranata, Sumarna. *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes*. Bandung: Remaja Rosdakarya, 2005.
- Thoha, M. Chabib. *Teknik Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada, 1996.