

BAB VI

SYARAT-SYARAT TEST YANG BAIK

Tujuan : Dapat memahami dan terampil melakukan teknik analisis syarat-syarat tes, yaitu : Validitas (kesalahan). reliabilitas (keterandalan). Usebilitas (kepraktisan).Objektifitas (objektif)

- a. Dapat memahami dan terampil melakukan tehnik analisis butir soal, yaitu : Analisis objektif. Analisis subjektif

Hasil pengukuran dan penelitian erat hubungannya dengan baik tidaknya teknik dan jenis serta alat penilaian/pengukuran yang dipergunakan, dalam hubungan ini misalnya alat penilaian/pengukuran test, haruslah memenuhi syarat-syarat yang baik. Untuk itu seorang guru haruslah memahami syarat-syarat tes yang baik dan dapat menganalisis kebaikan alat yang disusun. Dengan mengetahui tingkat kebaikan alat penilaian/pengukuran tersebut penasiran hasil dari penialian/pengukuran tersebut semakin komprehensif. Selain dari pada itu, hasil analisis dapat dipergunakan untuk memperbaiki alay yang dipergunakan yang akhirnya dapat didokumentasikan dalam rangka pembuatan bank soal yang memudahkan dan mengembangkan alat penilaian pada masa yang akan datang.

Berikut ini diutarakan syarat-syarat tes yang baik dan tehnik analisis syarat-syarat tersebut.

1. Relaiabilitas

1.1. Pengertian reliabilitas

Jika ditinjau dari sudut etimologis kata reliability berasal dari bahasa inggris yang asal katanya to rely yang artinya percaya (kepada), reliable artinya dapat dipercaya. Dari definisi para ahli dapat dilihat bahwa reliability uatu tes menunjukkan masalah ketetapan (consistency) dari

hasil tes tersebut, dengan demikian reliabilitas adalah syarat tes yang menyangkut kepercayaan kita tentang ketetapan hasil bila dilakukan dua atau lebih pada kelompok yang sama, contoh :

TABEL : 6.01 : HASIL SUATU TES DARI 5 ORANG MURID

No. urut	Nama	Hasil I		Hasil II	
1	Asi	50	3	80	2
2	Bili	45	4	70	3
3	Kani	60	1	90	1
4	Dini	30	5	50	5
5	Soni	57	2	60	4

Dari tabel di atas terlihat bahwa rank anak yang sama dari kedua hasil hanyalah dua orang sedang yang lain tidak, jadi ketetapan hasil kurang. Ketetapan yang dimaksud bukanlah skor tetapi mencakup ketetapan kedudukan anak dalam kelompoknya (ranking). Kalau tes yang dilaksanakan menunjukkan hal seperti di atas sedang aktor luar seperti cara pelaksanaan, kondisi dan situasi diantara pelaksanaan yang pertama dan kedua tidak berbeda (terdapat hanya sedikit penyimpangan maka dapatlah dikatakan tes yang bersangkutan kurang reliable (kurang dapat dipercaya).

Syarat reliabilitas sangat dituntut dalam tes yang standar misalnya tes inteligensi, minat dan lain-lain. Dalam tes hasil belajar yang disusun oleh guru pada umumnya kurang memperhatikan reliabilitas mengingat tes yang disusun hanya dipakai satu kali saja. Tetapi tes yang dipakai secara umum hendaknya mempunyai syarat reliabilitas mengingat tes yang bersangkutan mungkin dipergunakan pada masa yang akan datang.

1.2. Cara-cara menghitung reliabilitas

Tehnik ini disebut mencari reliabilitas dengan metode dua tes. Dalam hal ini dibuat dua perangkat soal

dimana bahan, bentuk, tingkat kesukaran item tes adalah sama, yng berbeda hanyalah kalimat item. Kedua tes tersebut disebut uqivalent, dan diberikan kepada kelompok yang sama secara berturut-turut dengan mencari korelasi antara kedua hasil yang diperoleh dapat dilihat tingkat reliabilitas te, koefisien yang diperoleh disebut : coefficient of equivalent atau correlation of equivalent.

1) Test-retest reliability

Tehnik ini adalah teknik satu tes. Tes retest berarti dua kali melaksanakan tes yang sama pada kelompok yang sama. Dalam rangka tehnik ini perlu dijaga bahwa pada pelaksanaan yang pertama murid tidak mengambil/menyalin soal. Tehnik ini disebut self correlation dan koefisien yang diperoleh disebut coefficient of stability.

TABEL 6.02 : PERHITUNGAN UNTUK PELAKSANAAN RELIABILITAS TES

No. urut	Skor item		XY	X ²	Y ²
	Ganjil (X)	Genap (Y)			
1	20	18	360	400	324
2	18	17	306	324	289
3	17	16	272	289	256
4	15	16	240	225	256
5	10	12	120	100	144
6	17	14	238	289	196
7	14	10	140	196	100
8	13	10	130	169	100
9	15	12	180	225	144
10	8	9	72	64	81
11	10	8	80	100	64
12	4	5	20	16	25
13	7	4	28	49	16
14	9	8	72	81	64
15	12	12	144	144	144
16	15	18	270	225	324
17	10	8	80	100	64
18	11	13	143	121	169
19	9	5	45	81	25
20	7	5	35	49	25
21	8	7	56	64	49
22	16	14	224	256	196
23	17	15	255	289	225
24	15	12	180	225	144
25	6	10	60	36	100
26	7	8	56	49	64

27	5	7	35	25	49
28	1	9	99	121	81
29	21	18	378	441	324
30	13	10	130	169	100
Jumlah	360	330	4448	4922	4142

Dimana : r_{II} = Coefficient o Reliability
 r^{1I} = Coeiciant correlation

2) Split-half reliability

Metode aplit-half reliability adalah metode satu tes. Perangkat tes yang diberikan dianggap dua macam tes yang sejajar, jadi satu test dibagi dua (split-half) dalam memeriksa dan membuat skor murid. Pembagian yang dimaksud dilaksanakan dengan item bernomor genap dan item bernomor ganjil. Hal ini lebih baik karena biasanya perangkat tes tersusun dari item yang mudah menuju item yang sukar. Jadi dalam menskor jawaban dibuat dua set skor yaitu skor murid dalam tes yang bernomor ganjil dan skor murid dalam tes yang bernomor genap. Kedua skor ini bernomor ganjil dan skor murid dalam tes yang bernomor genap. Kedua skor ini dicari korelasinya dan koefisien yang diperoleh menunjukkan tingkat reliabilitasnya tes yang disebut coefficient of internal consistency atau split-halves correlation.

Metode ini dimajukan oleh Speramen Brown dengan rumus :

$$r_{II} = \frac{2 r^{1I}}{1 + r^{1I}}$$

Mungkin masih banyak metode lain untuk menghitung reliabilitas suatu tes namun perhitungan secara sederhana tehnik diuraikan di atas cukup memadai. Untuk lebih memudahkan pemahaman berikut ini diberikan contoh menghitung reliabilitas suatu tes.

Contoh yang dimaksud adalah dengan teknik split-half reliability, dengan langkah-langkah berikut :

- Tentukan skor anak untuk item bernomor ganjil dan item bernomor genap.

- Carilah korelasi kedua skor tersebut.
- Carilah koefisien reliabilitas dengan rumus Spearmen Brown.
- Mean skor item ganjil = $\frac{360}{30} = 12$ (Mx)
- Mean skor item genap = $\frac{330}{30} = 11$ (My)
- Jumlah perkalian masing-masing skor pada item ganjil = 4922 ($\sum X^2$)
- Jumlah kuadrat masing-masing skor item ganjil = 4141 ($\sum Y^2$)
- Jumlah peserta = 30 orang (N)

Korelasi antara skor item genap adalah :

$$r = \frac{\sum XY - N(Mx)(My)}{\sqrt{\sum x^2 - NMx^2}(\sum Y^2 - NMy^2)}$$

$$r = \frac{4448 - 3960}{\sqrt{(4922 - 4320)(4142 - 3630)}}$$

$$r = \frac{488}{\sqrt{602 \times 512}} = \frac{488}{\sqrt{308224}}$$

$$= \frac{488}{555,2} = 0,88$$

$$R = \frac{2r}{1+r} = \frac{2 \times 0,88}{1+0,88} = \frac{2 \times 0,88}{1,88} = \mathbf{0,94}$$

Bila diperhatikan besarnya koefisien korelasi yang diperoleh ini ((0,94) maka dapat dikatakan korelasi sangat tinggi, jadi : Tes Reliable.

Pembatasan kasar dari angka korelasi adalah sebagai berikut :

- 0,80 ke atas : korelasi sangat tinggi
- 0,50 s/d 0,79 : korelasi tinggi
- 0,30 s/d 0,49 : korelasi sedang
- 0,20 s/d 0,29 : korelasi rendah
- 0,00 s/d 0,19 : korelasi sangat rendah

3) Reliabilitas Equivalent

Teknik ini disebut juga mencari reliabilitas dengan metode dua tes. Dalam hal ini disusun dua set tes yang equivalent. Tes yang equivalent berarti bahwa bahan, bentuk dan tingkat kesukaran butir soal adalah sejajar/parallel yang berbeda hanyalah perumusan kalimat dari

setiap butir soal. Di samping itu tes yang equivalent dapat juga diartikan. Dua set tes yang mempunyai kriteria yang sama dan tingkat kesukaran yang sama. Kedua tes ini diberikan kepada kelompok yang sama secara berturut-turut (jarak pelaksanaan yang pertama dan kedua tidak terlalu lama, lebih kurang satu atau dua hari). Skor masing-masing anak dari kedua tes itu dihitung korelasinya, dengan rumus korelasi langsung. Angka korelasi yang diperoleh disebut coefficient of equivalent atau correlation of equivalent. Semakin besar koefisien korelasi yang diperoleh itu semakin tinggi pula reliabilitas dari tes tersebut.

4) Inter-item Consistency

Metode ini dimaksudkan untuk mencari reliabilitas dengan memakai consistency dari setiap butir soal. Hal ini dikembangkan oleh Kuder dan Richardson yaitu dengan cara mengkorelasikan skor-skor item dengan skor total keseluruhan tes. Indeks reliabilitas yang didapat melalui cara ini sama dengan yang didapat metode perubahan yaitu koefisien konsistensi internal. Contoh berikut ini dapat memperjelas metode tersebut : Misalkan ada perangkat tes dengan sepuluh butir soal, yang dilaksanakan kepada sepuluh orang murid. Untuk jawaban murid yang benar diberi soal 1 (satu) dan jawaban murid yang salah diberi skor 0 (nol) sehingga diperoleh tabel berikut :

TABEL 6.03: HASIL PEMERIKSAAN PERANGKAT TES YANG DILAKUKAN DALAM HUBUNGAN DENGAN PERHITUNGAN INTER-ITEM CONSISENCY

No urut	Nama	NOMOR SOAL										X
		1	2	3	4	5	6	7	8	9	10	
1	Andi	1	0	1	1	1	0	1	1	0	1	7
2	Bobby	1	1	1	1	1	1	1	1	1	1	10
3	Corry	1	1	0	1	1	1	1	1	1	1	9
4	Dono	0	1	0	1	1	1	1	1	1	1	8
5	Eva	1	1	0	1	1	1	0	1	0	1	7
6	Fuad	1	1	1	0	1	1	0	1	1	0	7
7	Ganda	0	1	1	0	1	1	0	1	1	1	7
8	Henry	1	1	1	0	0	1	1	1	1	1	8
9	Irma	1	1	1	1	1	0	1	1	1	1	9
10	joni	1	1	0	1	1	1	1	1	1	1	8
	RA	8	9	6	7	9	8	7	10	8	8	80
	P	0,8	0,9	0,6	0,7	0,9	0,8	0,7	1,0	0,8	0	,8
	Q	0,2	0,1	0,4	0,3	0,1	0,2	0,3	0,0	0,2	0	,2
	PQ	0,16	0,9	0,24	0,21	0,9	0,16	0,21	0,0	0,16	0	3
											0	,1
											6	,2

Berdasarkan data diatas dapat ditentukan koefisien reliabilitas internal sebagai berikut :

Rumus yang dibuat oleh Kuder Richardson adalah :
Keterangan :

R_a = koefisien reliabilitas yang dicari
 n = jumlah butir soal
 S_{t^2} = varians skor total
 P = Proporsi siswa yang menjawab sebuah butir soal dengan benar
 Q = $1 - p$

Dengan kegunaan rumus di atas dapat dicari koefisien reliabilitas dari data tabel 6.03 sebagai berikut :

$$S_{t^2} = 3,82$$

$$PQ = 3,12$$

$$\begin{aligned}
 r_{tt} &= \left(\frac{n}{n-1} \right) \left(\frac{st^2 - PQ}{st^2} \right) = \left(\frac{10}{10-1} \right) \left(\frac{3,82-3,12}{3,82} \right) \\
 &= \left(\frac{10}{10-1} \right) \left(\frac{0,7}{3,82} \right) \\
 &= 1,11 \times 0,18 \\
 &= 0,1998 \\
 &= \underline{0,20}
 \end{aligned}$$

Dari angka koefisien tersebut (0,20) berarti reliabilitas internal tes tersebut rendah, jadi kurang reliable. Hal-hal yang Mempengaruhi Reliabilitas

1) Ada dua masalah pokok yang mempengaruhi reliabilitas tes, yaitu :

1. Masalah yang bersangkutan dengan perangkat tes.

a. Panjang-pendeknya kalimat item tes.

Item yang kalimatnya terlalu pendek apalagi angkat mudah diingat akan mengurangi reliabilitas tes karena setelah pelaksanaan yang pertama murid akan membahasnya dan bila dilaksanakan untuk kedua kalinya semua murid telah dapat menjawab dengan benar. Demikian juga item yang kalimatnya terlalu panjang akan menimbulkan keraguan dalam member jawaban atau murid tidak berusaha menjawabnya.

b. Tingka kesukaran tes/perangkat tes.

Item yang sangat sukar dan sangat mudah akan mengurangi tingkat reliabilitasnya te, karena

item yang sangat sukar akan selalu salah dijawab oleh kemungkinan seluruh kelas sukar akan selalu dijawab oleh kemungkinan seluruh kelas sebaiknya item yang sangat mudah akan selalu benar dijawab oleh kemungkinan seluruh murid. Sehingga bila dalam perangkat tes terdapat banyak item yang kesukarannya ekstrim akan mengurangi variasi skor dengan demikian akan mengurangi tingkat reliabilitas.

c. Kesesuaian waktu pelaksanaan tes.

Jumlah waktu yang dipergunakan hendaknya disesuaikan dengan jumlah item dan tingkat kesukaran item. Waktu yang “agak lama”, dimana setengah atau lebih dari murid yang dites sudah selesai sebelum waktu yang ditentukan habis, akan mengurangi sifat reliabilitas apalagi kalau metode perhitungannya melalui test-retest.

Seseorang anak yang sudah menyelesaikan pekerjaannya sepuluh menit sebelum waktu tes habis pada pelaksanaan pertama, mungkin akan menghabiskan waktu yang tersedia pada pelaksanaan kedua, sehingga waktu yang dipergunakan tidak sejajar lagi untuk kedua pelaksanaan tes itu.

2. Masalah yang berhubungan dengan pelaksanaan tes.

Masalah yang berhubungan dengan pelaksanaan/penyajian tes adalah faktor yang sangat berpengaruh dalam reliabilitas tes, termasuk diantaranya adalah :

a. Pengawas penyajian tes.

Tes yang pengawas penyajiannya tidak baik akan mengakibatkan murid-murid saling bekerja sama, hal ini akan mempengaruhi variasi nilai. Apalagi kalau teknik penelitian yang digunakan adalah tes-retest. Perbedaan pengawasan pada pelaksanaan yang pertama dengan yang kedua akan banyak mempengaruhi nilai murid.

b. Kondisi dan situasi penyajian

Kondisi murid dan kondisi/situasi tempat pelaksanaan tes erat hubungannya dengan tingkat reliabilitas tes. Murid yang kurang keunggulannya bekerja akan menghasilkan nilai yang kurang reliabel.

- c. Butir soal yang sudah diketahui anak, jelas akan mengurangi tingkat reliabilitas tes. Hal inilah yang menyebabkan diadakannya kode etik tes psikologis.

2. Validitas

2.1. Pengertian validitas

Dari uraian para ahli validitas tes adalah hal sesuai tidaknya isi tes dengan kriteria/tujuan tes. Untuk memudahkan pengertian dapat dijelaskan sebagai berikut : kalau kita hendak mengukur tinggi seseorang maka ukuran yang cocok adalah meter, dan bila kita hendak mengukur berat sesuatu benda ukuran yang cocok adalah timbangan yang satuannya disebut gram/kilogram. Dengan demikian juga halnya dengan tes, apabila kita mau mengukur pengetahuan siap dari anak, maka pertanyaan berikut adalah tidak cocok (tidak valid).

- Mengapa tumbuhan yang tidak kena sinar matahari tidak subur ?
- Apakah akibat pengangguran ?

Demikian juga halnya tes psikologis yang disusun di suatu Negara tidak selalu cocok untuk dipakai di Negara lain, hal ini disebabkan latar belakang kebudayaan dari kedua Negara tersebut tidak sama, misalnya : pertanyaan “apakah apel itu?” di Negara barat adalah termasuk pertanyaan yang tingkat keukarannya rendah, sedang di Indonesia pertanyaan itu adalah termasuk sukar karena buah apel bukan buah-buahan yang (umum) dijumpai. Dengan demikian harus dimodifikasi misalnya dengan pertanyaan “Apakah pisang itu?”

Satu perangkat tes yang menghasilkan skor tidak berkorelasi dengan skor tes lain di mana tes lain itu diperkirakan sudah baik, maka tes yang bersangkutan dapat dikatakan valid. Memperhatikan ada beberapa

aktor yang mempengaruhi tes, maka ada bermacam-macam jenis validitas sesuai dengan factor dimakud. Berikut ini akan memperjelas macam-macam validitas.

2.2. Jenis-jenis validitas

1) Face validity

Face validity disebut juga validitas semu. Diebut demikian karena sebenarnya jenis ini tidak menunjukkan keesuaian hasil tes dengan kriteria, tetapi menyangkut “tampang” tes itu yaitu bentuk luar seperti bentuk penulisan item, kerta yang dipergunakan dan lain-lain. Memang jenis validitas ini nampaknya nilainya rendah karena bentuk tes yang menarik, dicetak rapid an kertas yang dipergunakan baik belum tentu isi tes itu valid. Namun demikian, validitas semu perlu mendapat perhatian. Tes yang bentuk luarnya ditata rapi membangkitkan suatu kesan bahwa tes itu baik sehingga dengan demikian peserta akan bersungguh-sungguh dalam bekerja. Di sekolah para guru sering terikat akan masalah efisiensi sehingga mengurangi efektivitas.

2) Conten validity

Logical validity sering disebut conten validity dan ada yang menyebut curricular validity menyangkut perumusan item-item dan kesesuaian isi/makna item-item dengan tujuan pengukuran. Perumusan kalimat-kalimat dalam perangkat tes harus memenuhi tata bahasa yang baik, serta mudah dipahami. Di sekolah masalah validitas kurikulum yakni kesesuaian perangkat tes dengan isi kurikulum harusdiperhatikan. Lebih tegas dikatakan bahwa setiap penilaian hasil belajar harus mengacu kepada isi kurikulum yang berlaku.

Bagi guru pengertian ini harus mencakup pelaksanaan pembelajaran sehubungan dengan

kurikulum tersebut. Perangkat tes yang disusun/akan dilaksanakan guru bisa tidak memenuhi curricular validity. Walaupun perangkat tes yang disusunnya mengacu kepada isi kurikulum namun ada beberapa pokok bahasan yang belum diajarkan turut diujikan.

Di samping itu validitas ini juga menunjukkan kepada kemampuan perangkat tes untuk menggambarkan kelakuan yang diteliti dan sampai dimana hubungan antara hasil yang digambarkannya dengan kelakuan yang lain. Jadi, tes hendaknya mempunyai item-item yang dapat mengungkapkan semua tingkah laku sebagai manifestasi dari faktor yang diteliti. Suatu perangkat tes yang bertujuan untuk mengukur kemampuan murid-murid dalam menganalisa sesuatu pokok masalah, maka item-item yang disusun harus dapat mengungkapkan manifestasi dari kemampuan membandingkan, memberi kritik, menunjukkan baik buruknya sesuatu dan lain-lain. taksonomi Bloom sehubungan dengan kata-kata operasional dalam tujuan pembelajaran.

3) Predictive validity

Predictive validity (validitas ramal) bermakna tara keserasian suatu tes dalam meramalkan tingkah laku (keberhasilan) murid yang dites pada masa yang akan datang. Seseorang murid yang berhasil dalam mencapai angka tertinggi dalam tes Matematika yang dibuat oleh guru A akan memperoleh angka tertinggi pula pada tes yang dibuat oleh guru B apabila te yang dibuat oleh keua guru tersebut mempunyai predictive validity yang tinggi.

Dalam tes seleksi yang masuk sesuatu sekolah hendaknya mempunyai predictive validity yang tinggi. Dengan kata lain seseorang yang lulus dalam suatu seleksi, akan sukses mengikuti program yang diberikan bila alat seleksi yang dipergunakan mempunyai predictive validity yang tinggi. Kedua validity belakangan ini (predictive validity dan

factorial validity) sangat dituntut oleh tes yang standar seperti tes psikologis.

3. Objektivitas

3.1. Pengertian objektivitas

Alice Crow, menguraikan masalah objektivitas sebagai berikut :“The Administrator must follow exact procedure in administering the test, scoring responses, and interpreting results. Otherwise, data obtained through its administration by different persons, lack employing his own method of procedure, are not comparable... ..these direction must be understood clearly and without any deviations”.

Uraian ini menunjukkan bahwa setiap test itu haruslah dikonstruksikan dengan pertimbangan yang semaksimal-maksimalnya. Haruslah jelas apa yang ditetapkan sebagai jawaban dari item-item test dengan bentuk item mana item itu cocok. Haru pula ditetapkan lebih dahulu procedure di dalam manila jawaban-jawaban dan bagaimana pula menganalisa hasil-hasil tersebut. Setelah itu dilaksanakan/ditetapkan maka seseorang pelaksana atau pembuat test haruslah mengikuti prosedur-prosedure yang telah ditetapkan itu dengan tidak mengikutsertakan pertimbangan lain lagi yang sama sekali tidak berhubungan dengan test yang dibuat.

Masalah yang disebutkan di atas itu terutama berlaku kepada bentuk test yang informal, yaitu test yang belum distandard. Test yang distandard umumnya telah mempunyai pedoman jawaban yang benar. Cara scoring dan mempunyai norma untuk interpretasi. Namun demikian perlu juga dijaga bahwa test yang distandard pun mungkin terjadi penyimpangan dalam hal scoring. Petugas scoring yang kurang memahami rumus scoring yang telah ditetapkan atau kurang bersiat jujur karena satu dan lain hal akan mengurangi sifat objektivitas ini.

Berikut ini masalah objektivitas itu akan dibahas lebih luas.

3.2. Faktor-faktor yang berhubungan dengan syarat objektivitas

Seperti yang disebutkan di atas masalah objektivitas, secara garis besarnya menyangkut masalah :

a. Pengkonstruksian test

Test informal sering dikonstruksikan tidak memenuhi langkah-langkah yang seharusnya dilaksanakan. Item-item adalah dibuat saja. Dalam rangka achievement test sering seseorang membuat test hanya berdasarkan bahan pelajaran apa yang diingatnya saja, tidak memakai buku sumber yang sekaligus merupakan buku wajib/pegangan dalam mata pelajaran sebenarnya masih kurang baik, mengingat bahwa yang akan di test itu meliputi skill dan attitude dari murid. Dengan hanya mempedomani buku itu saja maka terungkaplah hanya masalah penguasaan murid terhadap bahan pelajaran yang diberikan (knowledge).

Oleh karena itu buku wajib/pegangan guru, serta tujuan instruksional khusus dari setiap pelajaran, merupakan pedoman yang harus digunakan dalam mengkonstruksi pelajaran, merupakan pedoman yang harus digunakan dalam mengkonstruksikan item-item.

b. Penentuan jawaban

Masalah yang kedua ini penting ditekankan dalam bentuk test essay. Guru harus lebih dahulu menetapkan jawaban-jawaban yang dituntut dari setiap soal yang akan ditanyakan. Dalam penentuan jawaban-jawaban yang dianggap benar ini, harus pula ditentukan luasnya/dalamnya jawaban yang diminta. Mungkin saja jawaban yang diminta lebih dari satu macam. Untuk itu harus ditentukan terlebih dahulu, urutan pentingnya atau urutan tepatnya jawaban yang bermacam-macam itu.

Karena masalah ini berhubungan erat dengan bentuk dan susunan kalimat dari setiap soal. Salah satu hal yang perlu diingat ialah menghindari kata-kata yang meragukan. Objektipun mungkin sering terjadi apabila kalimat/ kemungkinan jawaban yang dicantumkan kurang tepat/meragukan.

c. Penentuan nilai.

Masalah penentuan nilai sangat erat hubungannya dengan penentuan jawaban. Yang pertama harus ditentukan ialah bentuk nilai yang diberikan, apakah nilai dalam bentuk angka atau nilai dalam bentuk rating. Nilai bentuk angka sebaiknya diberi jarak 0 sampai dengan 100 atau 0 sampai dengan 10.

Yang perlu mendapat perhatian dalam penentuan nilai ini adalah tingkat kesukaran item-item. Maksudnya bahwa besarnya nilai bergantung kepada tingkat kesukarannya setiap item. Bila hal ini telah didistribusikan hendaknya direncanakan juga pendistribusian nilai untuk setiap kemungkinan jawaban.

Benarnya jawaban dalam essay secara umum dapat dibagi tiga macam yaitu : benar, mendekati benar dan salah – hal ini terutama bagi soal yang menghendaki penguraian. Dalam soal bidang studi ilmu pasti harus dipertimbangkan pendistribusian nilai untuk proses perhitungan yang salah tetapi jawaban akhir benar dan proses perhitungan yang benar tetapi jawaban akhir salah.

Dengan adanya penentuan nilai untuk jawaban – jawaban yang telah ditentukan, hal ini menghindari terjadinya pengaruh dari sipenilai. Dan perlu ditambahkan di sini, seorang penilai jangan menggantungkan besarnya nilai dengan kebersihan/ baiknya tulisan, kecuali hal ini penting memang merupakan kriteria test yang dilaksanakan

misalnya dalam menulis indah atau dalam menggambar.

Untuk menghindari suatu subjektivitas penilaian, hendaknya para penilai harus lebih dahulu mengkoreksi diri sendiri apakah memang pada saat ia memeriksa itu kondisi badan dan psikisnya baik.

Penelitian-penelitian yang dilaksanakan membuktikan bahwa penilaian suatu hasil test (scoring) hendaknya dilaksanakan pada pagi hari di mana keadaan kondisi fisik makin baik. gangguan terhadap suasana waktu menilai hendaknya dihindarkan. Waktu menilai menuntut suatu ketenangan, dan harus dilaksanakan pada waktu-waktu khusus dan tertentu. Jangan sampai terjadi seseorang menilai pekerjaan murid sambil ia mengerjakan pekerjaan lain. Dengan singkat dapat disimpulkan bahwa waktu, suasana, kondisi serta situasi seaktu pemeriksaan pekerjaan murid harus diciptakan sebaik mungkin sehingga pemeriksaan benar-benar memusatkan perhatian kepada pekerjaan tersebut.

Bila test itu merupakan test yang dilaksanakan oleh suatu panitia maka sebaiknya test penilaian harus diadakan.

Dengan melakukan cara : tiga orang memeriksa masing-masing untuk satu kerja pekerjaan murid, adalah sangat baik. Nilai akhir adalah nilai rata-rata untuk ketiga orang pemeriksa, dengan catatan bahwa perbedaan ketiga nilai yang diberikan tidak terlalu besar (pada EBTANAS SMTP/SMTA pada masa-masa yang lain, perbedaan itu paling maksimum 2 point dalam jarak nilai 0 s/d 19). Bila terjadi perbedaan nilai yang terlalu menyolok maka ketiga nilai itu dibatalkan dan pekerjaan anak dinilai ulang.

d. Analisa Nilai

Dalam buku-buku evaluasi analisa nilai disebut juga interpretasi *score*. Seperti yang disebutkan di atas bahwa interpretasi *score* ini dalam test kemampuan yang distandard tidak mengalami kesulitan lagi, karena normal interpretasi, dan walaupun ada norma, harus ditinjau latar belakang dari tes itu. Untuk latar belakang yang sangat berbeda perlu adanya peninjauan terhadap norma yang dibuat dalam manual tes itu.

Dengan mempergunakan teknik-teknik interpretasi *score* akan lebih baik dan tepat. Analisa nilai erat hubungannya dengan jenis test, system analisa pendidikan yang dianut, ungsi test dan keadaan umum dari yang di test. Penjelasan mengenai ini akan diuraikan pada bab berikut ini.

4. Teknik analisis butir soal

Analisis objektif

Pemeriksaan, pemberian nilai dan penafsiran hasil tes berhubungan dengan penilaian keberhasilan siswa dalam mengikuti pembelajaran yang telah dilalui. Tinggi rendahnya nilai yang diperoleh murid dipengaruhi oleh banyak faktor. Salah satu diantaranya adalah perangkat tes. Pada bab terdahulu telah dibicarakan tentang analisis perangkat tes, mencakup syarat validitas dan reliabilitas. Validitas dan reliabilitas korelasi menunjuk kepada valid tidaknya tes atau reliable tidaknya tes yang bersangkutan. Adanya variasi koeisien menunjuk pula adanya item yang kurang baik dan ada item yang sudah memadai. Guru perlu menganalisa butir soal yang kurang baik sehingga berangsur-angsur dapat didokumentasikan sejumlah soal yang relative lebih bermutu, yang selanjutnya dapat digunakan pada masa yang akan datang. Analisis objektif dimaksudkan adalah pemeriksaan butir soal sehubungan dengan teknik perhitungan tertentu secara eksak. Perhitungan

dimaksud adalah perhitungan daya pembeda dan tingkat kesukaran item.

1) Daya pembeda (discriminating power)

Daya pembeda adalah angka yang menunjukkan koefisien korelasi jumlah yang menjawab benar dari kelompok murid yang pandai dengan jumlah menjawab benar dari kelompok murid yang lemah.

Berikut ini salah satu contoh perhitungan daya pembeda. Contoh perhitungan ini adalah salah satu teknik perhitungan dari semacam teknik yang dapat digunakan.

Langkah-langkah untuk menghitung koefisien korelasi adalah sebagai berikut :

pertama, Susunlah lembar jawaban dimulai dari lembar jawaban yang memperoleh skor paling tinggi sampai kepada lembar jawaban yang memperoleh skor paling rendah, *kedua*, Tentukan kelompok murid yang pandai berdasarkan skor kelompok murid yang lemah masing-masing 27 % dari seluruh peserta, jadi 27 % dari atas susunan adalah kelompok pandai dari 27 % dari bawah susunan adalah kelompok lemah, sisa lembar jawaban (40%) disisihkan dari perhitungan. Penentuan 27 % menurut ahli Kelley (1933) adalah sampel yang cukup representative. *ketiga* Selanjutnya dibuat tabel distribusi jawaban masing-masing untuk kelompok pandai (disingkat H) dan untuk kelompok lemah (disingkat L), lihat contoh tabel dibawah ini :

TABEL 6.04 : DISTRIBUSI JAWABAN OLEH KELOMPOK PANDAI

No. urut	Nama Murid	Nomor Butir Soal						
		1	2	3	4	5	6	dst
1	Andilala	√	√	√	x	√	√	dst
2	Sindang	x	-	√	√	x	√	dst
3								
4								
5								
6								
dst								
Jumlah jawaban betul								
Persentase jawaban betul								

Tanda √ = jawaban betul
 X = Jawaban salah
 - = tidak dikerjakan

Selanjutnya dibuat tabel untuk kelompok lemah.

- i. Setelah dilakukan pendaftaran jawaban setiap murid pada setiap nomor, selanjutnya dijumlahkan jawaban betul untu setiap item dalam setiap kelompok dan seterusnya dihitung persentase jawaban betul untuk masing-masing nomor soal dengan memakai rumus :

$$P_H = \frac{R_H}{N_H} \times 100\%$$
 = Persentase jawaban betul untuk kelompok pandai (percentage high)
 R_H = Jumlah jawaban betul oleh kelompok pandai (right high)
 N_H = Jumlah semua murid dalam kelompok pandai = 72 % dari seluruh peserta (Number Hight)

Rumus diatas adalah untuk menghitung persentase pada kelompok pandai sedang pada kelompok lemah, prinsip rumus adalah sama yakni :

$$P_L = \frac{R_L}{N_L} \times 100 \%$$

P_L = percentage lower (persentase jawaban betul oleh kelompok lemah)

R_L = Right lower (jumlah jawaban betul oleh kelompok lemah)

N_L = Number lower (jumlah murid dalam kelompok lemah = 27 % dari seluruh peserta).

- ii. Hasil perhitungan langkah keempat dipakai sebagai dasar perhitungan pada format analisis butir soal. Format analisis butir soal sekaligus untuk keempat menuliskan tingkat kesukaran soal.

TABEL 6.05 : FORMAT ANALISIS BUTIR SOAL

No. Soal	Kelompok pandai		Kelompok lemah		Daya pembeda	Tingkat kesukaran
	R_H	P_H	R_L	P_L		
1						
2						
3						
4						
5						
6						
Dst						

Untuk menghitung daya pembeda yaitu koefisien korelasi dipergunakan tabel flanagan.

Misalkan kita menemukan hal berikut pada tabel distribusi jawaban :

- Jumlah menjawab betul butir soal nomor (1) dari kelompok pandai adalah 7 orang ($R_H = 7$)

- Jumlah menjawab betul butir soal nomor (1) dari kelompok lemah adalah 2 orang ($R_L = 2$)
- Jumlah semua murid dari kelompok pandai (27 % dari seluruh murid yang dites adalah 14 orang ($N_H = 14$))
- Jumlah semua murid dari kelompok lemah (27 % dari seluruh murid yang dites adalah 14 orang ($N_L = 14$)).

Dari data ini akan dapat dicari persentase untuk setiap kelompok yaitu:

$$P_H = \frac{R_H}{N_H} \times 100\% = \frac{7}{14} \times 100\% = 50\%$$

$$P_L = \frac{R_L}{N_L} \times 100\% = \frac{2}{14} \times 100\% = 14,2\%$$

Dalam mempergunakan tabel flanagan, kita cari angka 50 angka pada lajur teratas (proportion of correct responses in the upper 27 percent), kemudian kita telusuri ke bawah sampai pada lajur yang segaris dengan angka 14 yang terdapat pada kolom tegak lurus (proportion of correct responses in the lower 27 percent), untuk butir soal nomor (1) ini ditentukan angka 42, dengan demikian koefisien korelasi item nomor (1) adalah 0,42.

Contoh yang lain adalah :

Item nomor (2) mempunyai data sebagai berikut :

$$R_H = 4 ; N_H = 12$$

$$R_L = 6 ; N_L = 12$$

Perhitungan persentase item nomor (2) adalah :

$$P_H = \frac{R_H}{N_H} \times 100$$

$$= \frac{4}{12} \times 100\% = 33,3\%$$

$$P_L = \frac{R_L}{N_L} \times 100 =$$

$$= \frac{6}{12} \times 100\% = 50\%$$

Karena P_H lebih dari P_L maka koefisien korelasinya adalah negative dan cara menemukan koefisien korelasi item nomor (2) dilakukan dengan cara berikut :

Angka persentase dari kelompok pandai (P_H) kita pakai pada angka tegak lurus di sebelah kiri sedang angka persentase dari kelompok lemah (P_L), kita pakai lajur angka yang terdapat di tabel.

Untuk data di atas (butir soal-soal) kita temukan angka dalam tabel sebagai perpotongan garis horizontal angka 33 dan garis vertical angka 50 adalah angka 17, jadi koefisien korelasi dari item (2) adalah $-0,17$.

2) Tingkat kesukaran (Difficulty level)

Tingkat kesukaran butir soal tes ditentukan oleh semua peserta yang memberi jawaban betul pada kedua kelompok pandai dan lemah. Rumus yang dipergunakan untuk menentukan tingkat kesukaran soal adalah :

$$P = \frac{R_H + R_L}{N_H + N_L} \times 100 \%$$

P = Tingkat kesukaran

N_H, N_L, R_H dan R_L Sama maksudnya dengan penjelasan yang terdapat di muka. Bila rumu tersebut dipergunakan dengan data yang terdapat pada conoth di atas di mana :

$$\text{Item nomor (1) : } R_H = 7 ; R_L = 2$$

$$N_H = 14 ; N_L = 14$$

$$\text{Item nomor (2) : } R_H = 4 ; R_L = 6$$

$$N_H = 12 ; N_L = 12$$

Maka :

- Tingkat kesukaran item nomor (1) adalah :

$$\begin{aligned} P &= \frac{R_H + R_L}{N_H + N_L} \times 100 \% \\ &= \frac{7 + 2}{14 + 14} \times 100 \% = 41,7 \% \end{aligned}$$

3) Interpretasi Hasil Analisa Butir Soal

Setelah kita memperoleh angka daya pembeda dan persentase tingkat kesukaran selanjutnya angka tersebut kita pakai menentuka apakah item-item perangkat soal

yang dibuat adalah baik atau tidak (memenuhi kriteria daya pembeda dan tingkat kesukaran yang baik). Biasanya terdapat korelasi yang positif (tinggi) antara angka daya pembeda dan angka tingkat kesukaran. Untuk dapat menarik kesimpulan tentang kedua angka tersebut ditetapkan rasio yang dipedomani adalah sebagai berikut :

- a. `sebuah soal tes dikatakan mempunyai kualitas tinggi jika butir soal tersebut mampu membedakan kelompok murid yang pandai dengan kelompok murid lemah.
- b. Butir soal yang baik apabila jumlah peserta dari kelompok pandai relative lebih banyak menjawab betul butir soal tersebut dibandingkan dengan kelompok murid yang lemah.
- c. Terdapat korelasi yang positif antara daya pembeda dan tingkat kesukaran. Butir soal yang terlalu sukar tidak akan dapat membedakan kelompok murid pandai dengan kelompok murid lemah karena butir soal itu tidak terjawab oleh kelompok murid pandai maupun kelompok murid lemah dan sebaliknya butir soal yang terlalu mudah tidak memiliki daya pembeda yang baik karena butir soal tersebut dapat dijawab dengan betul oleh semua peserta ujian.

Sehubungan dengan hal di atas berikut ini diberikan ketentuan sehubungan dengan interpretasi dari hasil perhitungan daya pembeda dan angka tingkat kesukaran. Sebelumnya perlu diingat bahwa angka tingkat kesukaran dari perhitungan di atas menunjukkan hal berikut : semakin besar angka tingkat kesukaran hal ini menggambarkan semakin mudah butir soal sebaliknya semakin kecil angka tingkat kesukaran hal ini menggambarkan semakin sukar butir soal yang bersangkutan. Lihat rumus, hal ini terjadi karena persentase tingkat kesukaran dihitung berdasarkan jumlah peserta yang menjawab betul setiap butir soal dan setiap kelompok.

Berlainan halnya dengan angka daya pembeda, semakin tinggi mutu daya pembeda dan semakin kecil angka daya pembeda semakin rendah mutu daya pembeda butir soal tersebut.

Ketentuan dalam mempergunakan angka daya pembeda dan tingkat kesukaran adalah :

1) Kriteria pemilihan butir soal

Sebuah butir soal yang baik (memadai) sebagai soal tes yang dapat dipergunakan apabila memenuhi persyaratan berikut :

Koefisien korelasi daya pembeda paling rendah 0,20

2) Persentase kesukaran paling rendah 10 % dan paling tinggi 90 %

Penggolongan tingkat kesukaran butir soal dapat dilakukan sebagai berikut :

30 % - 20 % = butir soal sukar

70 % - 31 % = butir soal sedang

90 % - 71 % = butir soal mudah

3) Butir soal yang tidak memenuhi ketentuan poin(1) perlu direvisi kembali bila ingin dipergunakan lagi.

Sehubungan dengan ini dari tabel format analisis butir soal dapat juga dilihat berikut ini :

a. Daya pembeda yang bertanda negative atau 0,00 (nol) adalah suatu pertanda bahwa butir soal yang bersangkutan “meragukan” peserta. Murid dari kelompok yang pandai tidak akan mengerjakan butir soal yang bersangkutan, sedang murid dari kelompok lemah yang mempunyai kecenderungan untuk “mencoba-coba” akan mengerjakannya, yang kemungkinan jawaban coba-coba itu betul. Hal ini akan membedakan angka daya pembeda yang negative.

b. Daya pembeda yang kurang dari 0,20 dan tingkat kesukaran lebih kecil dari 10 % hal ini menunjuk butir soal yang bersangkutan terlalu sukar dan apabila tingkat kesukaran lebih dari 90% hal ini berarti butir soal terlalu mudah.

Setelah kita mendapat data tentang daya pembeda dan tingkat kesukaran, diperlukan analisis subjektif terutama bagi butir soal yang tidak memenuhi kriteria untuk memudahkan kita dalam merevisi butir soal tersebut.

Analisis Subjektif

Analisis subjektif bermakna meneliti perumusan butir soal berdasarkan akal sehat (common sense) . jadi dilihat dari logika berfikir terhadap ide yang dikandung oleh setiap butir tes, apakah sudah tepat atau perlu diperbaiki perumusan kalimat-kalimatnya. Analisis subjektif dapat dilakukan sebelum tes dilaksanakan atau dalam rangka merevisi soal berdasarkan hasil analisis objektif. Untuk analisis subjektif dapat dipedomani hal-hal berikut :

1) Menilai Perangkat Tes

Di dalam menilai suatu soal tes sebagai suatu keseluruhan perlu dipertimbangkan apakah :

- a. Soal tes yang bersangkutan menguji pengetahuan dan suatu keterampilan yang cukup tinggi.
- b. Ada cara lain yang lebih baik untuk menguji tujuan yang dimaksud.
- c. Soal tes tersebut mempunyai tingkat kesukaran yang memadai bagi populasi yang akan diuji.

Tiga pertanyaan di atas harus diperunakan dalam menganalisis butir soal. Pertanyaan (a) menekankan kepada rating tidaknya masalah yang dikandung oleh setiap butir tes dilihat dari kriteria tes yang ditetapkan. Kesalahan yang sering terjadi adalah guru menyusun hanya bergantung kepada sumber pelajaran tanpa selalu memperhatikan tujuan pembelajaran yang akan dicapai. Pertanyaan (b) bermakna bahwa penggunaan bentuk tes harus dilihat dari tujuan. Harus diingat bahwa dalam penilaian/pengukuran masih ada teknik non tes, ada bermacam-macam bentuk tes, ada bermacam-macam cara pelaksanaan tes. Hal ini menghunjuk perlunya dipertimbangan cara yang tepat untuk meneliti suatu

tujuan. Dalam konteks tujuan pengajaran juga banyak variasi (lihat Taksonomi Bloom). Pertanyaan (c) menyangkut tingkat kesukaran. Tingkat kesukaran “memadai” bermakna bahwa perangkat tes yang bersangkutan dapat dikerjakan dengan betul sesuai tingkat kemampuan murid, hal ini bukan berarti semua murid dapat menjawab setiap pertanyaan dengan betul, tetapi distribusi skor tes menunjuk kurva normal. Kesukaran suatu item tergantung dari kesukaran materi dan ketersembunyiannya kunci jawaban yang dirancang. Guru harus memahami bahwa ukuran ukra tidaknya harus dilihat dari sudut murid, bukan dari sudut guru, jadi titik tolak tingkat kesukaran adalah murid. Menilai batang tes.

Di dalam menilai batang (stem) soal tes, harus dipertimbangkan apakah batangan soal tes yang dimaksud itu :

- a. Telah dirumuskan masalah secara jelas (unambigous)
- b. Berisi informasi yang berlebihan
- c. Bisa dirumuskan secara lebih jelas lagi.

Seperti yang dijelaskan pada petunjuk konstruksi tes, satu butir soal bisa meragukan pengertian apabila kalimatnya terlalu panjang tau terlalu pendek. Pemilihan kata yang lebih tepat adalah hal yang sukar. Demikian juga hanya tentang kata-kata yang sifatnya subjektif. Informasi yang berlebihan dapat menjadi kunci dalam memberi jawaban atau dapat membingungkan murid karena mungkin antara informasi yang satu terdapat perbedaan dengan informasi yang lain atau malah terdapat dua atau lebih informasi yang bertentangan. Harus diingat bahwa (stem) soal bukan pokok dari soal tes tetapi pengantar dalam menghunjuk masalah yang dimajukan. Sesuatu masalah dapat dirumuskan dengan bermacam-macam kalimat yang mempunyai

makna sama atau hampir bersamaan. Untuk itu harus dipilih perumusan yang paling tepat dan tegas.

2) Menilai kemungkinan jawaban (options)

Di dalam menilai kemungkinan jawaban (options), harus dipertimbangkan apakah pilihan :

- a. cukup parallel strukturnya dan atau homogeny isinya
- b. sesuai dengan batang soal tes dari segi logika/tata bahasa
- c. bisa dirumuskan lebih jelas dan atau parallel lagi
- d. ada yang memenuhi syarat sebagai kunci jawaban pokok uji yang bersangkutan.

Semua kemungkinan jawaban dari setiap soal harus mengandung makna ruang lingkup yang sama. Bila pokok uji misalnya tentang jenis ikan yang diperkembangkan dengan perkolaman, maka alternatif jawaban hendaknya adalah jenis ikan bukan jenis lain, lihat contoh berikut : Jenis ikan yang dikembangkan dengan perkolaman: a) ikan laut, b) udang, c) ikan mas, c) belut.

Strukturanya yang dimaksud dalam butir (a) adalah perumusan yaitu tentang keseragaman tentang panjang/ pendeknya kalimat dari setiap options. Sesuatu options yang dirumuskan dengan kalimat yang panjang sedang options lainnya dirumuskan dengan kalimat pendek menarik peserta untuk memilih options dengan kalimat yang lebih panjang tersebut. Hal-hal seperti ini memberi isyarat dalam memilih kunci jawaban. Kesalahan yang sering terjadi, ialah adanya keraguan hubungan antara batang (stem) soal dengan kemungkinan jawaban.

Contoh :

Tumbuhan yang telah mati, rusak atau tidak menghasilkan lagi segera :

- a. ditebang dan dibakar
- b. dimusnahkan untuk pupuk
- c. diremajakan dengan tanaman baru
- d. diremajakan dengan dan lain-lain

contoh tersebut menunjukkan adanya keraguan antara batang tes dengan options. Pada batang tes tidak disebut “tumbuhan yang mana”, padahal maksud penulis mungkin adalah tumbuhan yang dipelihara karena ada perkataan tidak menghasilkan namun hal itu belum tentu dapat dilihat oleh peserta ujian.

3) Menilai Kunci Jawaban

Di dalam menilai kunci jawaban, harus dipertimbangkan apakah :

- a. Ada pilihan lain yang juga memenuhi syarat sebagai kunci jawaban
- b. Kunci jawaban benar-benar menjawab masalah yang dikemukakan batang soal tes.
- c. Kunci jawaban bisa dirumuskan secara lebih “tersembunyi” bila dirumuskan aslinya terlalu “jelas”.

Kunci jawaban sangat berhubungan dengan batang tes. Seperti dijelaskan terdahulu bahwa option harus paralel dalam struktur dan homogen isinya, dengan demikian batang tes tidak jelas dan tegas maka akan mengakibatkan kunci jawaban kabur, atau terdapat dua atau lebih options yang dapat dipilih menjadi kunci jawaban.

Untuk jelasnya dapat dilihat dalam contoh berikut :a) rakit penyeberangan,b) jembatan,c) jembatan gantung,d.) jembatan darurat.

Jawaban yang benar (kunci jawaban) dari soal di atas bisa saja keempat-empatnya.

4) Menilai Pengecoh (Distractors)

Di dalam menilai pengecoh (distractor) harus dipertimbangkan apakah pengecoh :

- a. Ada yang dapat dipertimbangkan sebagai kunci jawaban juga

- b. Cukup “mauk akal” sehingga pengikut ujian yang kurang menguasai bahan dapat terkecoh olehnya.
- c. Malah memberi isyarat kearah kunci jawaban, sehingga pengikut berhasil menjawab dengan benar sebagai akibat dari membaca pengecoh saja.

Yang dimaksudkan dengan pengecoh (distrators) adalah daya tarik dari setiap kemungkinan jawaban untuk dipilih menjadi kunci jawaban . jadi setiap kemungkinan jawaban haruslah “nampaknya” bisa dipilih sebagai kunci jawaban terutama bagi murid lemah/ sedang. Di dalam melaksanakan analisis subjektif, hendaknya dilakukan oleh “teman sejawat” yang mempunyai latar belakang yang sama, biasanya orang lain lebih dapat melihat kekurangan dari perumusan butir soal tersbut. Analisis subjektif ini adalah sebagai pelengkap bukan merupakan penentu, yang lebih dapat dipercaya adalah analisis objektif.

Untuk melihat efektivitas pengecoh ini dapat juga dilakukan dengan cara melihat perbedaan jumlah peerta yang memilih dari kelompok pandai dan dari kelompok lemah sebagai jawaban yang benar. Berikut ini diberikan suatu contoh :

TABEL 6.06 : DISTRIBUSI PILIHAN JAWABAN SISWA PADA BUTIR SOAL

Nomor Soal : 1						Kunci
Jawaban : C						
Pilihan jawaban Jumlah yang memilih dari:	A	B	C	D	E	Tidak Menjawab
Kelompok pandai	14	41	77	7	0	1
Kelompok lemah	9	35	66	66	0	10
Jumlah	23	76	143	143	0	11

Interpretasi dari item di atas dapat ditarik kesimpulan sebagai berikut :

1. Distraktor kemungkinan jawaban C dan D merupakan distraktor yang baik karena dipilih oleh kelompok pandai lebih banyak dari kelompok lemah (C = Kunci jawaban). Sedang kemungkinan jawaban D lebih banyak dipilih oleh kelompok lemah dibandingkan dengan kelompok pandai (D = Bukan kunci jawaban).
2. Distraktor kemungkinan jawaban A dan B tidak mempunyai distraktor yang baik karena kelompok pandai lebih banyak memilihnya dibandingkan dengan kelompok pandai (D = bukan kunci jawaban).
3. Distraktor kemungkinan jawaban E tidak efektif karena tidak dapat menarik jawaban seorang pun dari kelompok pandai maupun dari kelompok lemah.
4. Bila misalnya kunci jawaban adalah B, pada contoh diatas maka distraktor kemungkinan jawaban C disebut distraktor yang menyesatkan, karena lebih banyak dipilih peserta dibandingkan dengan kunci jawaban.

UNIMED

THE
Character Building
UNIVERSITY

Rangkuman :

1. Syarat validitas tes menyangkut ketetapan perangkat tes untuk mengukur tujuan tes, ketetapan perumusan butir soal dengan aspek dan materi yang akan diuji.
2. Syarat reliabilitas tes menyangkut ketetapan penggambaran hasil dari tes, yang hal ini berhubungan dengan kebaikan perumusan butir soal dan situasi/kondisi pelaksanaan tes tersebut.
3. Syarat kepraktisan tes menyangkut biaya, waktu, fasilitas yang digunakan, suasana dan kebudayaan. Hal ini berarti pelaksanaan tes, penyiapan tes, pelaksanaan tes terutama dalam pemeriksaan dan interpretasi hasil tes.
4. Syarat keobjektifan tes menyangkut baik tidaknya penyiapan tes, pelaksanaan tes terutama dalam pemeriksaan dan interpretasi hasil tes.
5. Agar hasil penilaian/pengukuran dapat mencapai tujuan sesuai yang ditetapkan syarat validitas, reliabilitas, kepraktisan dan objektivitas harus benar-benar diperhitungkan sebaik-baiknya.
6. Agar alat penilaian/tes selalu dapat diperkembangkan sehingga pada suatu waktu diperoleh bank soal, hendaknya dilakukan analisis validitas, reliabilitas dan butir soal, terutama pada perangkat tes yang ditunjukkan dalam menilai hasil belajar pada catur ulan atau semester.
 1. prioritas I (paling penting) menuju prioritas V (yang kurang penting)!