

CIRI-CIRI TES YANG BAIK DAN
CARA MENENTUKANNYA

OLEH

A R L I S

JURUSAN PENDIDIKAN BIOLOGI

FPMIPA IKIP PADANG

MILIK PERPUSTAKAAN IKIP PADANG

TERIMA

5-7-95

SUMBER/HARGA

bi

KOLEKSI

KRI

NO INVENTARIS

1313/ku/95-C(2)

NO SERI/KASI

371.26 arl e1

MILIK UPT PERPUSTAKAAN
IKIP PADANG

Kata Pengantar

Sesuatu pekerjaan belum tahu kita berhasil atau tidak sebelum diukur/diuji/dievaluasi. Demikian juga kita di dalam dunia pendidikan, terutama di dalam proses belajar mengajar, kita belum tahu apakah kita berhasil mengajar atau belum sebelum objek yang diajar (siswa) kita uji/di tes apa yang telah mereka kuasai.

Untuk mengetahui hasil suatu pekerjaan itu sesuai dengan apa yang kita kehendaki tentu juga diukur/diuji dengan alat yang tepat/sesuai dengan apa yang harus kita ukur/uji. Umpamanya untuk mengukur berat sesuatu tentu kita pakai timbangan. Kalau timbangan kita kurang baik tentu hasil pengukuran kita tidak akan menunjukkan berat yang sebenarnya. Hal ini sering kita dengar suatu barang beratnya 1 kg, tetapi setelah ditimbang dengan timbangan yang baik (benar), berat barang tersebut tidak tepat 1 kg, umpamanya 0,9 kg atau 1,05 kg dan lain sebagainya. Dengan demikian kalau kita akan mengukur sesuatu pakailah alat yang baik/benar sesuai dengan apa yang diukur. Untuk mengetahui alat yang baik/benar itu terlebih dahulu harus ditara atau dikalibrasi. Demikian juga alat yang akan dipakai untuk mengukur hasil belajar yang kita sebut dengan tes/alat evaluasi hasil belajar. Untuk mendapatkan alat evaluasi hasil belajar yang baik harus ditara/dikalibrasi terlebih dahulu yang dalam hal ini kita sebut analisis item. Inilah yang akan kita jelaskan dalam tulisan ini yaitu apa ciri tes yang baik dan bagaimana cara menentukannya.

Mudah-mudahan tulisan ini bermanfaat bagi kita semua, amin.

Daftar Isi

Kata Pengantar	i
I. Penadshuluan	1
II. Cara Menentukan Tes Yang Baik	4
III. Contoh Cara Menentukan Tes Yang Baik	12
IV. Daftar Bacaan	27

I. PENDAHULUAN

Dalam pengukuran pendidikan ada tiga istilah yang perlu diperhatikan yaitu tes, pengukuran dan penilaian. Dalam pengalaman kita sehari-hari terlihat bahwa tes itu adalah soal-soal yang diberikan kepada murid atau anak didik untuk mereka jawab atau selesaikan. Dari jawaban yang mereka berikan kita atau guru akan memberikan suatu tanda atau angka yang disebut hasil mereka. Hasil ini tentu sesuai dengan tujuan kita memberi tes. Jadi tes merupakan suatu alat untuk mengetahui apa yang ingin diketahui. Raka Joni (1986) menyatakan bahwa secara operasional tes bisa didefinisikan sebagai sejumlah tugas yang harus dikerjakan oleh orang yang di tes (Testee). Kesanggupan dalam menunaikan tugas-tugas ini dapat dinyatakan dalam bentuk angka-angka. Angka yang diberikan ini biasanya disebut prestasi yang di tes tersebut dalam hal/bidang yang dites. Kalau murid dites dalam penguasaan materi pelajaran yang diwajibkan, maka hasil yang diberikan disebut prestasi belajar murid tersebut. Jadi dengan tes kita dapat mengukur sejauh mana murid tersebut telah menguasai materi yang diberikan. Ini sesuai dengan apa yang dikemukakan oleh Raka Joni (1986) bahwa tes adalah salah satu alat untuk mengadakan pengukuran.

Kalau yang ingin diukur itu perestasi belajar siswa, maka tesnya disebut tes prestasi belajar. Data atau angka yang diperoleh dari hasil pengukuran ini diterjemah sesuai dengan tujuan pengukuran tersebut umpamanya testi dinyatakan baik, kurang baik, kurang dan sebagainya atau pintar, kurang pintar atau lulus, tidak lulus dan sebagainya. Jadi dari data hasil pengukuran melalui suatu proses pertimbangan dengan memakai patokan tertentu kita menetapkan suatu keputusan. Pekerjaan beginilah yang disebut penilaian. Sehubungan dengan penilaian ini Raka Joni (1986) menyatakan bahwa penilaian bisa digambarkan sebagai suatu proses dimana kita mempertimbangkan sesuatu barang atau gejala dengan mempergunakan patokan tertentu. Patokan tersebut mengandung pengertian baik, tidak baik, memadai, tidak memadai, memenuhi syarat, tidak memenuhi syarat dan sebagainya. Dengan perkataan lain kita mengadakan value jugment.

Dari apa yang dibicarakan di atas terlihat bahwa tes adalah sebagai salah satu alat pengukuran pendidikan terutama prestasi belajar yang menghasilkan informasi kuantitatif yang dipakai sebagai dasar untuk mengarahkan penilaian. Sudah barang tentu kecermatan pelukisan data/informasi sangat tergantung kepada kecermatan/kuualitas alat atau tes yang dipakai.

Di atas dikatakan bahwa tes itu adalah berupa alat. Dalam penilaian pendidikan biasanya dibicarakan

dua jenis tes yaitu tes buatan guru dan tes yang sudah distandardisir. Suatu tes dinamakan terstandardisir apabila soal-soalnya telah diujicobakan dalam keadaan yang sama atau hampir sama dengan keadaan untuk mana tes itu disusun sehingga pada tes tersebut telah terdapat kalibrasi dalam bentuk norma-norma (Raka Joni 1986). Tes buatan guru, sebagaimana namanya adalah tes yang dibuat oleh guru sendiri untuk memenuhi kebutuhan sendiri.

Berbicara mengenai sifat tes, kita melihat dari sudut pemberian nilai dibedakan atas 2 macam yaitu tes subjektif dan objektif. Pada tes subjektif pemberi skor banyak memakai pertimbangan subjektif sedangkan pada tes objektif pemberi skor boleh dikatakan tidak usah mempertimbangkan kebenaran dari jawaban testi sebab pemeriksa hanya mencocokkan saja jawaban dari testi dengan kunci jawaban yang telah disediakan. Masing-masing tes di atas mempunyai keunggulan dan kelemahan sendiri-sendiri (Raka Joni, 1986).

Di atas, di kemukakan bahwa untuk mendapatkan informasi/data yang baik (dapat dipercaya) sangat ditentukan salah satunya oleh alat ukur yang baik. Dalam penilaian pendidikan, untuk mendapatkan informasi/data yang baik sehubungan dengan prestasi belajar anak tentu dibutuhkan alat ukur/tes prestasi belajar yang baik juga. Bagaimana caranya untuk mendapatkan tes yang baik tersebut ikutilah penjelasan halaman berikut.

II. CARA MENENTUKAN TES YANG BAIK

Wayan Nurkencana dkk (1983) menyatakan bahwa baik buruknya suatu tes atau alat evaluasi dapat ditinjau dari validitas, reliabilitas, tingkat kesukaran dan daya beda. Sedangkan menurut Suharsini Arikunto (1989) ciri-ciri tes yang baik yaitu memiliki validitas, objektivitas dan ekonomis. Tetapi Raka Joni (1986) mengemukakan bahwa ciri-ciri yang harus dimiliki oleh suatu tes agar bisa menunaikan fungsinya sebagaimana yang diharapkan adalah validitas, reliabilitas, tingkatan kesukaran, kemampuan diskriminasi dan syarat-syarat tambahan lainnya seperti keluasan ruang lingkup, kemudahan administrasi serta scoring dan lain-lain. Disamping apa yang disampaikan di atas tentang ciri-ciri tes/soal yang baik, Direktorat Jenderal Pendidikan Dasar dan Menengah (1985) menyampaikan bahwa untuk mempertinggi ketepatan pengukuran perlu diperhatikan kaidah-kaidah diantaranya :

- a. Pokok soal (stem) yang merupakan permasalahan harus dirumuskan secara jelas.
- b. Perumusan pokok soal dan alternatif jawaban hendaknya merupakan pernyataan yang diperlukan saja.
- c. Untuk setiap soal hanya ada satu jawaban yang benar atau yang paling benar.
- d. Pada pokok soal (stem) sedapat mungkin dicegah perumusan pernyataan yang bersifat negatif.
- e. Alternatif jawaban (option) sebaiknya logis dan pengecoh harus berfungsi (menarik).

- f. Diusahakan agar tidak ada petunjuk untuk jawaban yang benar.
- g. Diusahakan agar alternatif jawaban (option) homogen, baik dari segi isi/materi maupun panjang pendeknya pernyataan.

Dari uraian di atas tentu dipertanyakan apakah yang dimaksud dengan masing-masing ciri tersebut dan bagaimana menentukan suatu tes telah memiliki ciri yang dikatakan baik. Mengenai hal ini Raka Joni (1986) mengemukakan bahwa suatu tes dikatakan valid apabila ia mengukur apa yang mau diukur dengan kemungkinan yang sekecil mungkin dipengaruhi oleh faktor-faktor yang tidak relevant dan Suharsini Arikunto (1989) mengatakan bahwa sebuah tes disebut valid apabila tes itu dapat tepat mengukur apa yang hendak diukur.

Kapankah suatu tes reliabel atau dikatakan mempunyai reliabilitas yang baik. Dalam hal ini Raka Joni (1986) mengemukakan bahwa suatu tes dikatakan reliabel apabila hasil pengukurannya mantap. Kemantapan ini dapat dilihat dari tiga segi yaitu (a) sebagai kemantapan hasil mengukur ulangan dengan tes yang sama yang menghasilkan indeks stabilitas, (b) sebagai kemantapan hasil mengukur dengan dua buah tes yang paralel, yang dianggap sama yang menghasilkan indeks ekuivalensi kesamaan dan (c) kemantapan hasil mengukur dari masing-masing soal dihubungkan dengan kemantapan tes secara keseluruhan yang menghasilkan indeks konsistensi internal atau keman-

tapan internal. Suharsini Arianto (1989) mengatakan bahwa realibilitas berhubungan dengan masalah kepercayaan. Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Maka pengertian reliabilitas tes, berhubungan dengan masalah ketepatan hasil tes. Dari penjelasan ini tentu pasti kita dapat mencari realibilitas suatu tes atau kita dapat menetapkan reliabilitas suatu tes. Untuk keperluan ini Wayan Nurkancana dkk (1983) mengemukakan bahwa taraf reliabilitas suatu tes dapat dicari dengan (a) teknik ulang, (b) teknik bentuk paralel, (c) teknik belah dua. Kemudian untuk mencari reliabilitas seluruh tes dipergunakan rumus Spearman Brown :

$$r_n = \frac{N \cdot r_{12}}{1 + (N-1) r_{12}}$$

dimana : r_n = Koefisien korelasi seluruh tes.

N = perbandingan antara panjang tes seluruhnya dengan tes yang dikorelasikan.

r_{12} = Koefisien korelasi antara sebagian tes dengan bagian tes yang lainnya.

Menurut Slameto (1988) dalam mencari ukuran reliabilitas, maka yang dikorelasikan itu ialah sepasang rangkaian skor yang diperoleh dengan cara yang disebut "Cara belah dua" dan dipakai rumus korelasi product moment Pearson.

Untuk mencari reliabilitas keseluruhan soal, dicari koefisien korelasi dengan menggunakan rumus Spearman Brown yaitu :

$$r_1 = \frac{2 r_s}{1 + r_s}$$

dimana : r_1 = koefisien korelasi keseluruhan soal.

r_s = koefisien korelasi separoh soal yang diketemukan dengan membagi dua keseluruhan soal.

Dibahagian muka disampaikan bahwa suatu tes selain memperhatikan validitas dan reliabilitas perlu juga dilihat derajat kesukaran dan daya bedanya.

Wayan Nurkancana dkk (1983) mengemukakan bahwa suatu tes tidak boleh terlalu mudah dan juga tidak boleh terlalu sukar. Jadi item yang baik adalah item yang mempunyai derajat kesukaran tertentu yaitu bergerak antara 25 % sampai 75 %. Untuk menentukan derajat kesukaran item ini dipakai rumus :

$$DK = \frac{WL + nH}{nL + nH} \times 100 \%$$

dimana : DK = derajat kesukaran

nL = jumlah individu kelompok bawah

nH = jumlah individu kelompok atas

WL = jumlah individu kelompok bawah yang salah

WH = jumlah individu kelompok atas yang salah

Lebih lanjut Wayan Nurkencana menyatakan, disamping derajat kesukaran, suatu item harus juga mempunyai daya beda tertentu sebab suatu tes yang baik adalah tes yang dapat membedakan murid yang pintar dengan murid yang bodoh. Untuk menentukan daya beda suatu item dipakai rumus :

$$DB = \frac{WL - WH}{n}$$

dimana : DB = daya beda

n = jumlah kelompok atas atau bawah.

Item yang dapat dipakai diharapkan item yang mempunyai daya beda 0,40 ke atas.

Setelah kita mengkaji masalah tes yang baik yaitu mempertimbangkan validitasnya, reliabilitas, derajat kesukaran dan daya beda tes tersebut, maka sebaiknya juga dipertimbangkan tingkatan perkembangan kognitif apa yang harus diukur oleh tes tersebut. Dalam masalah ini Amirudin Arief (1976) mengemukakan, supaya dalam merumuskan atau menyusun alat evaluasi (tes) supaya memperhitungkan prosentase pembobotan hasil belajarnya (tingkat kelasnya) dan pembobotan itu dinyatakan dengan perbandingan prosentase antara :

1. Ungkapan fakta (Recall fact)
2. Pemahaman (Comprehension)
3. Penerapan/Aplikasi (aplikation).

Bagi siswa tahun pertama ungkapan fakta lebih banyak, pemahaman secukupnya dan aplikasi belum dipakai (diberikan). Bagi siswa tahun kedua, penggunaan ungkapan fakta sudah dikurangi, pemahaman bertambah dan aplikasi sudah dapat diberikan. Bagi siswa tahun ketiga ungkapan fakta bertambah sedikit, pemahaman cukup banyak dan aplikasi sudah bertambah lagi.

Jadi Terlihat bagi kita semakin tinggi tingkatan siswa yang di tes, semakin besar pula bobot soal kearah perkembangan kognitif yang lebih tinggi dan semakin sedikit pada tingkat kognitif rendah.

Sehubungan dengan prosentase pembobotan tes ini Direktorat Jenderal Pendidikan Dasar dan Menengah Depdikbud (1983) mengemukakan dalam pedoman penulisan soal EBTA-NAS 1983/1984, untuk tingkat SLTP berbanding seperti 29 : 48 : 23 antara ingatan, pemahaman dan aplikasi. Ini kalau dibulatkan menjadi 1 : 2 : 1.

Sesuai dengan pendapat di atas yaitu semakin tinggi jenjang/tingkat pendidikan anak semakin tinggi pula pembobotan tes ke arah perkembangan yang lebih tinggi, maka pada tingkatan SMA sudah seharusnya ada tes yang berupa analisa, sintesa dan evaluasi. Dalam hal ini diperkirakan prosentase pembobotannya antara ingatan, pemahaman, aplikasi dan lain-lain (analisa, sintesa, evaluasi) seperti 1 : 2 : 2 : 1.

Pada bagian muka disampaikan bahwa menurut Direktorat

Jendral Pendidikan Dasar dan Menengah, dalam penulisan soal perlu juga diperhatikan kaidah-kaidah yang berlaku, diantaranya adalah supaya alternatif jawab (option) sebaiknya logis dan pengecoh harus berfungsi.

Sehubungan dengan hal di atas Wayan Nurkancana (1983, hal 136) mengemukakan bahwa salah satu hal yang perlu diperhatikan dalam mengadakan revisi item ialah efektifitas dari pada masing-masing option yang digunakan dalam item tersebut. Untuk mengetahui apakah suatu option berfungsi secara efektif atau tidak, ditempuh prosedur sebagai berikut :

- a. Ambil 27 % lembar jawaban yang mendapat skor tertinggi dan 27 % lembar jawaban yang mendapat skor terendah (Jadi sama dengan prosedur mencari derajat kesukaran dan daya beda).
- b. Buat tabel sejumlah item yang akan diuji efektivitas option-optionnya, sebagai berikut :

Option	a	b	c	d	e
Kelompok					
Atas					
Bawah					

- c. Isikan distribusi pilihan terhadap option yang disediakan baik untuk kelompok atas maupun untuk kelompok bawah.

d. Berdasarkan distribusi pilihan kelompok atas dan kelompok bawah, maka dapat dihitung option mana yang berfungsi secara efektif dan option mana yang tidak berfungsi secara efektif. Pedoman yang digunakan untuk menentukan efektivitas suatu option adalah sebagai berikut :

1). Untuk option kunci.

- Jumlah pemilih kelompok atas dan kelompok bawah tidak kurang dari 25% tetapi tidak lebih dari 75%.
- Frekuensi pilihan kelompok atas harus lebih tinggi dari pada frekuensi pilihan kelompok bawah.

2). Untuk option pengecoh (distractor).

- Jumlah pemilih kelompok atas dan kelompok bawah, minimal adalah 25% kali satu per dua kali jumlah option pengecoh kali jumlah kelompok atas ditambah kelompok bawah.
- Frekuensi pilihan kelompok bawah harus lebih tinggi dari pada frekuensi pilihan kelompok atas.

Demikianlah ciri-ciri tes yang baik dan cara menentukannya.

III. CONTOH CARA MENENTUKAN TES YANG BAIK

Pada bahagian muka tela h disampaikan sehubungan dengan ciri tes yang baik dan cara menentukannya secara teoritis. Untuk lebih memantapkan teoritis tersebut sebaiknya diberikan contoh dan nanti dicoba mengerjakannya. Untuk keperluan contoh ini kami mengambil satu set tes/alat evaluasi hasil belajar mata pelajaran biologi dari satu daerah yang terdiri dari 65 item tes dengan jumlah peserta tes (testi) sebanyak 265 orang.

Sebagaimana disampaikan pada bahagian terdahulu hasil analisa berupa komposisi item tes, derajat kesukaran, daya beda item berfungsinya option dari masing-masing item, reliabilitas tes, dan kaidah penulisannya.

Setelah dilakukan analisa terhadap 65 item tes alat evaluasi hasil belajar yang dipakai sebagai contoh, diperoleh item yang bersifat ingatan 29 item, pemahaman 25 item, aplikasi 9 item, dan lain-lain 2 item. Dalam hal ini tidak terdapat item yang tergolong evaluasi. Kalau hasil pengelompokan item ini dijadikan prosentase akan berupa 44,62 : 38,46 : 13,85 : 3,08.

Proporsi yang diharapkan , seperti dinyatakan di muka adalah 1 : 2 : 2 : 1 . Kalau ini dijadikan prosentase akan berupa 16,67 : 33,33 : 33,33 : 16,67. Untuk melihat apakah proporsi di atas sesuai dengan yang diharapkan dilakukan perhitungan dengan

memakai rumus :

$$Z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Proporsi soal yang bersifat ingatan

$$\begin{aligned} p &= \frac{X_1 + X_2}{n_1 + n_2} \\ &= \frac{16,67 + 44,62}{100 + 100} \\ &= \frac{61,29}{200} = 0,306 \end{aligned}$$

$$\begin{aligned} q &= 1 - p \\ &= 1 - 0,306 = 0,694 \end{aligned}$$

$$\begin{aligned} Z &= \frac{x_1/n_1 - x_2/n_2}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \\ &= \frac{\frac{16,67}{100} - \frac{44,62}{100}}{\sqrt{0,306 \times 0,694 \left(\frac{1}{100} + \frac{1}{100} \right)}} \\ &= \frac{-0,2795}{\sqrt{0,212364 \times \frac{2}{100}}} = \frac{-0,2795}{0,0652} \end{aligned}$$

$$Z = -4,29$$

H diterima bila $-Z_{\frac{1}{2}}(1 - \alpha) < Z < Z_{\frac{1}{2}}(1 - \alpha)$.

$Z_{\frac{1}{2}}(1 - \alpha)$ untuk signifikansi 5% adalah 1,96.

Jadi H diterima jika $-1,96 < Z < 1,96$

Dengan demikian untuk proporsi soal yang bersifat ingatan belum sesuai dengan yang diharapkan, sebab

harga $Z = -4,29$ lebih kecil dari $-1,96$. Disini terlihat proporsi soal yang bersifat iangatan tersebut lebih besar dari yang diharapkan.

Proporsi soal yang bersifat pemahaman

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{33,33 + 38,46}{100 + 100}$$

$$= \frac{71,79}{200} = 0,3590$$

$$q = 1 - p = 1 - 0,3590$$

$$= 0,6410$$

$$Z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$= \frac{\frac{33,33}{100} - \frac{38,46}{100}}{\sqrt{0,3590 \times 0,6410 \left(\frac{1}{100} + \frac{1}{100} \right)}}$$

$$= \frac{\frac{33,33}{100} - \frac{38,46}{100}}{\sqrt{0,004602}} = \frac{-0,0513}{0,0678}$$

$$Z = \frac{-513}{678} = -757$$

Sesuai dengan ketentuan di atas maka proporsi soal yang bersifat pemahaman sesuai dengan yang diharapkan sebab harga $Z = -0,757$ berada dalam daerah penerimaan H yaitu $-0,757$ lebih besar dari $-1,96$ dan lebih kecil dari $+1,96$.

Proporsi soal yang bersifat aplikasi

$$\begin{aligned}
 p &= \frac{x_1 + x_2}{n_1 + n_2} \\
 &= \frac{33,33 + 13,85}{100 + 100} \\
 &= \frac{47,18}{200} = 0,2359
 \end{aligned}$$

$$\begin{aligned}
 q &= 1 - p = 1 - 0,2359 \\
 &= 0,7641
 \end{aligned}$$

$$\begin{aligned}
 Z &= \frac{x_1/n_1 - x_2/n_2}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \\
 &= \frac{\frac{33,33}{100} - \frac{13,85}{100}}{\sqrt{0,2359 \times 0,7641 \left(\frac{1}{100} + \frac{1}{100} \right)}} \\
 &= \frac{0,1948}{\sqrt{0,003605}} = \frac{0,1948}{0,06004} \\
 &= 3,24
 \end{aligned}$$

Sesuai dengan ketentuan di muka, maka proporsi soal yang bersifat aplikasi belum sesuai dengan yang diharapkan sebab harga Z yang diperoleh di luar, lebih besar dari 1,96. Disini proporsi soal yang bersifat aplikasi tersebut lebih kecil proporsinya dari pada yang diharapkan.

Proporsi soal yang bersifat lain-lain (Analisa, sintesa dan evaluasi).

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{16,67 + 3,08}{100 + 100} = \frac{19,75}{200}$$

$$= 0,9775$$

$$q = 1 - p = 1 - 0,9875$$

$$= 0,0125$$

$$Z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$= \frac{\frac{16,67}{100} - \frac{3,08}{100}}{\sqrt{0,9875 \times 0,0125 \left(\frac{1}{100} + \frac{1}{100} \right)}}$$

$$= \frac{\frac{13,59}{100}}{\frac{\sqrt{0,02468750}}{100}} = \frac{0,1359}{0,0157}$$

$$Z = 8,66$$

Seperti hal yang di atas tadi, proporsi soal yang bersifat lain-lain juga belum sesuai dengan yang diharapkan, sebab harga Z yang diperoleh di luar harga penerimaan H yaitu : 8,66 lebih besar dari 1,96.

Terlihat bahwa soal yang bersifat lain-lain tersebut jauh lebih kecil proporsinya dari pada yang diharapkan.

Dari semua yang dikemukakan di atas, proporsi alat evaluasi belajar yang dianalisa ini, yang bersifat pemahaman

sudah berada pada daerah yang diharapkan, sedangkan yang bersifat aplikasi lebih sedikit, demikian juga yang bersifat lain-lain (analisa, sintesa evaluasi) jauh lebih sedikit lagi dari yang diharapkan.

Sebaliknya soal yang bersifat ingatan lebih besar proporsinya dari pada yang dituntut dalam kurikulum.

Demikianlah hasil analisa proporsi dari alat evaluasi mata pelajaran biologi yang dipakai sebagai contoh.

untuk melihat daya beda (DB) dan derajat kesukaran (DK) dari tiap item tes dari alat evaluasi yang diteliti dilakukan seperti di bawah ini :

- Dari seluruh peserta tes diambil 27% sebagai kelompok bawah (kurang pandai).
- Menentukan berapa orang dari kelompok bawah yang menjawab item tersebut salah atau tidak menjawab sama sekali dan ini disimbulkan dengan WL.
- mentabulasi berapa orang dari kelompok atas yang menjawab item tersebut salah atau tidak menjawab sama sekali dan ini disimbulkan dengan WH.
- selanjutnya mencari WL + WH dan WL + WH
- untuk dapat daya beda yangkai rumus :

$$DB = \frac{WL - WH}{n}$$

- Untuk mendapatkan derajat kesukaran dipakai rumus :

$$DK = \frac{WL + WH}{n_L + n_H} \times 100\%$$

- Hasil perhitungan yang telah dilakukan adalah seperti tabel I; berikut ini :

TABEL II
DERAJAT KESUKARAN DAN DAYA BEDA
MASING-MASING ITEM TES

nomor item	WH	WL	WL + WH	WL - WH	D K	D B
1	6	26	3232	20	-22 %	0,28
2	59	63	122	4	85 %	0,06
3	33	39	72	6	50 %	0,08
4	23	53	76	30	53 %	0,42
5	33	39	72	6	50 %	0,08
6	37	31	68	-6	47 %	0,08
7	20	29	49	9	34 %	0,13
8	32	45	77	13	53 %	0,18
9	51	68	119	17	83 %	0,24
10	46	54	100	8	69 %	0,11
11	46	55	101	9	70 %	0,13
12	35	56	91	21	63 %	0,29
13	20	38	58	18	40 %	0,25
14	27	53	80	26	56 %	0,36
15	45	54	99	9	69 %	0,13
16	38	444	82	6	57 %	0,08
17	32	45	77	13	53 %	0,18
18	47	66	113	19	78 %	0,26
19	27	39	66	12	46 %	0,17
20	64	61	125	-3	87 %	0,04
21	11	55	66	44	46 %	0,61
22	40	51	91	11	63 %	0,15
23	2	10	12	8	-8 %	0,11
24	65	65	130	0	90 %	0
25	37	50	87	13	60 %	0,18
26	6	22	28	16	-19 %	0,22
27	7	32	39	25	27 %	0,35
28	68	72	140	4	97 %	0,06
29	47	64	111	17	77 %	0,24
30	59	63	122	4	85 %	0,06
31	44	59	103	15	72 %	0,21
32	67	65	132	-2	92 %	0,03
33	54	53	107	-1	74 %	0,01

Hal berikutnya yang sebaiknya dilihat untuk mengetahui kualitas alat evaluasi hasil belajar adalah reliabilitasnya. Hal ini dikerjakan dengan metatabulasi nilai peserta ujian (265 orang). dari hasil tabulasi ini diperoleh :

- rang nilai peserta dengan nomor soal ganjil adalah antara $23,5 - 6,5 = 17$

- rang nilai peserta dengan nomor soal genap adalah antara $19,5 - 5,5 = 14$

- interval untuk nilai peserta nomor soal ganjil yang memenuhi syarat antara $\frac{17}{7} = 2 \frac{3}{7}$ dan $\frac{17}{15} = 1 \frac{2}{15}$. Untuk mudahnya perhitungan dipakai interval 2.

- interval untuk nilai peserta nomor soal genap antara $\frac{14}{7} = 2$ dan $\frac{14}{15}$. Untuk memudahkan perhitungan dipakai interval 2.

Dengan memakai apa yang dikemukakan di atas dan masing-masing kelas interval didapatlah peta korelasi seperti berikut (TABEL II) (Sudjana, 1975, hal.262 - 264)

tabel II:

PETA KORELASI

x \ y	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	f	x'	fx'	fx ²	f.x'-y'
22-23			3	1	1	3	1		9	4	36	144	28
20-21				1	6	3	2		12	3	36	108	54
18-19		1	3	6	7	4	2	2	25	2	50	100	48
16-17	1	2	16	17	16	7	2	3	64	1	64	64	25
14-15	2	10	17	22	13	7	2	1	74	0	-0	-0	0
12-13		3	7	12	9	5	1		37	-1	-37	37	-9
10-11	1	5	8	10	3	2	2	1	32	-2	-64	144	8
8-9	1	3	1	2	1				8	-3	-24	72	27
6-7	1	1	2						4	-4	-16	64	28
f	6	25	57	71	56	31	12	7	265		45	733	209
y	-3	-2	-1	0	1	2	3	4					
/y	-18	-50	-57	0	56	62	36	28	57				
/y ²	9	100	57	0	56	124	108	112	611				
/x'y'	24	44	0	0	34	54	33	20	209				

Keterangan: x interval nilai siswa dengan nomor soal ganjil

y interval nilai siswa dengan nomor soal genap

$$M_x = \frac{\sum fx}{N} = \frac{45}{265} = 0,170$$

$$M_y = \frac{\sum fy}{N} = \frac{57}{265} = 0,215$$

$$SD_x = \sqrt{\frac{\sum fx^2}{N} - \frac{(\sum fx)^2}{N}}$$

$$= \sqrt{\frac{733}{265} - (0,170)^2}$$

$$= \sqrt{2,766 - 0,0289} = \sqrt{2,737} = 1,654$$

$$\begin{aligned}
 SD.y &= \sqrt{\frac{\sum fy'^2}{N} - \frac{(\sum fy')^2}{N}} \\
 &= \sqrt{\frac{611}{265} - \frac{(57)^2}{(265)}} \\
 &= \sqrt{2,305 - 0,046225} = 1,503
 \end{aligned}$$

koefisien korelasi product moment dicari dengan rumus

$$r = \frac{\frac{\sum fx'y'}{N} - M_x' \cdot M_y''}{SD.x' \times SD.y'}$$

$fx'y'$ diperoleh dengan mengalikan setiap frekuensi dengan x' kali y' pada setiap kolom interval x .

demikian juga untuk kolom interval y

$fx'y'$ pada kolom interval y 5-6 adalah $1x-3x-1 + 2x-3x0 + 1x-3x-2 + 1x-3x-3 + 1x-3x-4 = -3+0+6+9+12 = 24$.

$fx'y'$ pada kolom interval y 7-8 adalah $1x2x-2 + 2x1x-2 + 10x0x-2 + 3x-1x-2 + 5x-2x-2 + 3x-3x-2 + 1x-4x-2 = -4-4+0+6+20+10+8 = 44$.

dengan cara yang sama $fx'y'$ pada kolom interval y 9-10 11-12, 13-14, 15-16, 17-18, 19-20 diperoleh seperti pada peta korelasi.

$fx'y'$ pada kolom interval x 6-7 adalah $1x-4x-3 + 1x-4x-2 + 2x-4x-1 = 28$.

$fx'y'$ pada kolom interval x 8-9 adalah $1x-3x-3 + 3x-3x-2 + 1x-3x-1 + 2x-3x0 + 1x3x1 = 27$.

dengan cara yang sama $fx'y'$ pada kolom interval x 10-11, 12-13, 14-15, 16-17, 18-19, 20-21, 22-23 diperoleh seperti pada peta korelasi.

$$\begin{aligned} r &= \frac{\frac{209}{265} - 0,170 \times 0,215}{1,654 \times 1,503} \\ &= \frac{0,789 - 0,035550}{2,486} \\ &= 0,3026. \end{aligned}$$

r tabel untuk $N = 265$ dan signifikansi 5% adalah 0,115.

Jadi disini terlihat r yang diperoleh lebih besar dari r tabel untuk $N=265$ dan signifikansi 5%. Dengan demikian maka korelasi yang diperoleh adalah signifikan jadi ada korelasi antara skor kedua belahan tes (soal) yang di analisa.

keefisien korelasi keseluruhan soal adalah :

$$\begin{aligned} r_r &= \frac{2 r_s}{1+r_s} \\ &= \frac{2 \times 0,3026}{1+0,3026} = 0,4646. \end{aligned}$$

berdasarkan kriteria tingkat korelasi yang dikemukakan oleh Slamuto (1988, hal 215) yaitu :

$0,80 \leq r < 1,00$	korelasi sangat tinggi
$0,60 \leq r < 0,80$	korelasi tinggi
$0,40 \leq r < 0,60$	korelasi sedang
$0,20 \leq r < 0,40$	korelasi rendah
$0,00 \leq r < 0,20$	korelasi sangat rendah

maka dapat dikatakan bahwa tingkat reliabilitas soal tersebut berada pada tingkat sedang. sekarang marilah uilihat pula apakah masing-masing option sudah berfungsi secara efektif. cara kerjanya sudah disampaikan pada tinjauan pustaka.

Hasil analisis sehubungan dengan berfungsinya masing-masing option dari alat evaluasi yang dimaksud dengan memakai kreteria seperti disampaikan pada bagian terdahulu terlihat pada TABEL IV berikut :

TABEL IV
FREKUENSI JAWABAN DAN BERFUNGINYA OPTION

No. item	KEL. ATAS					KEL. BAWAH					EFEKTIVITAS OPTION				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
1.	1	66	2	1	2	3	46	9	12	2	-	-	+	+	-
2.	18	8	18	13	15	17	17	18	9	11	-	+	-	-	-
3.	11	46	4	9	2	20	24	4	19	5	+	+	-	+	+
4.	6	7	49	6	4	10	15	19	18	12	+	+	+	+	+
5.	39	21	4	6	2	33	16	11	6	6	+	-	+	-	+
6.	0	1	30	34	2	3	5	41	21	2	-	+	-	-	-
7.	1	12	52	7	0	2	15	43	12	0	-	+	+	+	-
8.	2	25	40	5	0	6	27	27	10	2	+	+	+	+	+
9.	1	21	1	42	7	8	4	3	51	6	+	-	-	+	-
10.	3	26	9	12	22	9	18	10	10	25	+	+	+	-	+
11.	26	18	8	15	7	17	19	14	12	10	+	+	+	-	+
12.	24	37	6	2	3	19	16	15	9	13	-	+	+	+	+
13.	52	5	10	3	2	34	7	11	17	3	+	+	+	+	+
14.	4	11	6	45	6	12	14	17	19	10	+	+	+	+	+
15.	20	13	9	27	3	13	19	14	18	8	-	+	+	+	+
16.	9	34	8	8	13	8	28	20	6	10	-	+	+	-	-
17.	6	2	3	40	21	6	4	10	27	25	-	+	+	+	+
18.	5	27	6	9	25	13	15	9	29	6	+	-	+	+	-

Keterangan: tanda srip (-) menunjukkan nomor soal tersebut pada kaidah kolom itu sebaiknya diperbaiki.

- Kolom : 1. Pokok soal (stem) yang merupakan permasalahan harus dirumuskan secara jelas.
2. Perumusan pokok soal dan alternatif jawaban hendaknya merupakan pertanyaan yang diperlukan saja.
3. Untuk setiap soal hanya ada satu jawaban yang benar atau yang paling benar.
4. Pada pokok soal (stem), sedapat mungkin dicegah perumusan pernyataan yang bersifat negatif.
5. Alternatif jawaban (option) sebaiknya logis dan pengecoh harus berfungsi (menarik).
6. Diusahakan agar tidak ada "petunjuk" untuk jawaban yang benar.
7. Diusahakan untuk mencegah penggunaan option yang terakhir berbunyi "semua pilihan jawaban di atas salah" atau "semua pilihan jawaban di atas benar".
8. Diusahakan agar alternatif jawaban (option) homogen, baik dari segi isi/materi maupun panjang pendeknya pertanyaan.
9. Apabila alternatif jawaban (option) berbentuk angka, susunlah secara berurutan mulai dari angka yang terkecil di atas dan yang terbesar di bawah.
10. Di dalam pokok soal (stem), diusahakan tidak menggunakan ungkapan atau kata-kata yang bersifat tidak tentu, seperti: kebanyakan, seringkali, kadang-kadang dan yang sejenis.

DAFTAR BACAAN

- Arikunto, Suharsini, 1989, Dasar-Dasar Evaluasi Pendidikan, Jakarta, Bina Aksara.
- Arlis dkk, 1983, Studi Perbandingan Tentang Komposisi Tes Hasil Belajar Bidang Studi IPA Pada SMA Sekotamadya Padang, IKIP Padang.
- Arlis, 1990, Tinjauan Terhadap Kualitas Tes Prestasi Belajar Mata Pelajaran Biologi Program A₂ Semester Januari - Juni 1990 SMA Negeri Kodya Bukittinggi, IKIP Padang
- DirjenPDM, DEPDIKBUD, 1985, Buku Panduan Penulisan Soal Evaluasi Tahap Akhir Nasional, Jakarta.
- Joni, T. Raka, 1986, Pengukuran Dan Penilaian Pendidikan, Surabaya, Karya Anda.
- Nurkancana, Wayan dkk, 1983, Evaluasi Pendidikan, Surabaya, Usaha Nasional.
- Slamete, 1988, Evaluasi Pendidikan, Jakarta, Bina Aksara
- Sudjana, 1975, Metoda Statistika, Bandung, Tarsite.