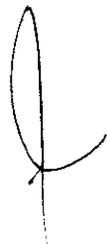


SYARAT-SYARAT PENYUSUNAN TEST HASIL BELAJAR

OLEH

DR. ALIASAR, M. Ed.

| | |
|----------------------|-------------------------|
| UNIVERSITAS PADJARAN | |
| NO. SURAT | 4. 7. 95 |
| DUMBUK HIMPUNAN | hi |
| KOLEKSI | VKI |
| NO. SURAT | 1287 / hi / 95 - 5. (2) |
| NO. SURAT | 371.26 ali 51 |



Disampaikan pada

PENATARAN DAN LOKAKARYA PEMANTAPAN SKS, EVALUASI BELAJAR, DAN BIMBINGAN AKADEMIK STAF PENGAJAR DI LINGKUNGAN KOPERATIS WILAYAH X, TANGGAL 25 NOVEMBER SAMPAI DENGAN 4 DESEMBER 1990, DI PADANG

DAFTAR ISI

| | |
|---------------------------------------|----|
| I. Latarbelakang | 1 |
| II. Validity (Ketepatan) | 2 |
| A. Content Validity | 3 |
| B. Construct Validity | 6 |
| C. Face Validity | 7 |
| D. Predictive Validity | 8 |
| E. Concurrent Validity | 8 |
| III. Reliability | 9 |
| A. Test Retest Method | 10 |
| B. Split- Half Method | 11 |
| C. Equivalent Forms | 14 |
| D. Kuder Richardson Formula (.KR.21.) | 15 |
| IV. Usability | 16 |
| A. Bahasa | 17 |
| B. Format Soal | 17 |
| C. Susunan Soal | 18 |
| V. Ekonomis Dalam Penyelenggaraan | 18 |
| VI. Kesimpulan | 19 |
| VII. References | 20 |

SYARAT-SYARAT PENYUSUNAN TEST HASIL BELAJAR

I. LATARBELAKANG / RASIONAL

Tidak dapat disangkal lagi bahwa seorang tenaga pengajar, disamping memberikan arah atau menyampaikan program belajar, juga dituntut untuk mengetahui keberhasilan warga belajar dalam mencapai tujuan belajar. Dalam proses belajar semua kegiatan guru, murid, serta petugas bimbingan dan administrasi diarahkan untuk menuju pencapaian tujuan program belajar. Kadang kala, proses belajar telah berlansung demikian lama, tetapi tidak diketahui apakah tujuan yang telah ditetapkan telah tercapai atau belum. Oleh sebab itu evaluasi merupakan suatu kegiatan yang esensial dalam setiap proses belajar, dan seorang guru atau tenaga instruktur tidak dapat berlepas diri dari pelaksanaan evaluasi itu.

Untuk melaksanakan suatu evaluasi hasil belajar, terlebih dulu harus dipersiapkan alat ukur yang tepat digunakan untuk memantau hasil belajar / tujuan pengajaran. Jika seseorang tenaga pengajar salah atau tidak mengetahui alat ukur yang tepat, besar kemungkinan pemantauan hasil belajar yang telah dilaksanakan akan mengalami kesalahan. Akibatnya, segala kesimpulan atau keputusan yang diambil berdasarkan hasil ukur yang salah akan berakibat fatal.

Dalam makalah yang sederhana ini akan diuraikan teori-teori atau pendapat para ahli dalam menyusun alat ukur, khususnya item test hasil belajar yang baik. Mudah-mudahan teman sejawat di Perguruan Tinggi, yang bukan dari bidang evaluasi ini, akan dapat faham menyusun test hasil belajar yang telah diprogramkannya.

Soal- soal ujian atau item test adalah alat ukur atau instrument untuk mengetahui hasil belajar. Instrument yang baik mempunyai syarat- syarat sebagaimana yang dikemukakan oleh Tuckman sebagai berikut:

1. Validity
2. Reliability
3. Usability ; (B. Tuckman, 1975, p.229 -302).

Selanjutnya S. Arikunto mengemukakan syarat- syarat utama dari suatu test yang baik adalah sebagai berikut:

1. Validitas
2. Reliabilitas
3. Objektivitas
4. Praktikabilitas
5. Ekonomis (S. Arikunto, 1990. hal.56).

Malaupun ke dua pendapat diatas kelihatannya terdapat perbedaan, namun pada dasarnya tidaklah ada perbedaan yang sangat esensial. Selanjutnya pada makalah ini akan dibahas satu persatu syarat- syarat dari instrument penilaian tersebut.

II. VALIDITY (KETEPATAN)

Apabila istilah validity ini dialih bahasakan ke dalam Bahasa Indonesia beberapa pengarang menggunakan istilah validitas. Istilah lain yang sering juga dipakai oleh pengarang kita yang sama artinya dengan valid ini adalah sehih; dengan demikian istilah validity diterjemahkan dengan kesahihan. Istilah sehih ini berasal dari bahasa Arab yang barangkali tidak 100% sama konsepnya dengan valid. Oleh karena itu dalam beberapa buku pengukuran dan penilaian hasil belajar yang berbahasa Indonesia masih banyak dipakai istilah aslinya validity yang juga banyak diterjemahkan dengan ketepatan.

Suatu alat ukur dikatakan valid apabila alat/ instrument tersebut sesungguhnya dapat mengukur objek yang hendak diukur. " Validity refers to the extent to which an instrument measures what it is intended to measure," (D. Ary et. al 1979, p. 196).

Istilah lain yang sering juga dipakai oleh pengarang Indonesia untuk pengganti valid adalah tepat. Jadi ketepatan adalah merupakan terjemahan untuk validity. Validity adalah kata benda, sedangkan valid kata keadaan atau sifat.

Lindvall, dan Nitko mengemukakan, " Validity, then, is the first consideration in the selection of an appropriate procedure- a procedure designed to permit the student to exhibit the behavior in the objective. " (Lindvall, and Nitko 1975, p. 22) .

Suatu hal yang sangat penting dalam penentuan validity dari suatu instrument adalah bagaimana prosedur menggunakan instrument tersebut. Instrument yang baik atau valid tidak akan menjadi valid kalau prosedur memakai/ menggunakannya salah (tak sesuai dengan ketentuan yang telah ditetapkan). Berikut ini akan diuraikan beberapa jenis validity.

A. Content Validity (Validitas Isi)

Suatu alat ukur / instrument test mempunyai content validity apabila keseluruhan dari items yang ditanyakan mewakili (representatif) terhadap materi kurikulum yang telah disajikan (diukur). Oleh karena materi yang ditanyakan harus terwakili di dalam soal-soal yang diberikan, maka sering juga disebut dengan curriculum validity .

Terwakili (representativeness) materi yang ditanyakan itu biasanya ditinjau dari dua sudut pandangan yaitu: 1) subjek atau

materi kurikulum, dan 2) taraf tingkatan ranah perubahan tingkah laku (lihat Taxonomy Bloom pada buku Zahara Idris dkk, 1979,hal.27).

Belajar atau learning pada prinsipnya adalah suatu proses yang dilaksanakan oleh si belajar dan sebagai akibatnya terjadilah perubahan tingkah laku di dalam diri si belajar tersebut. Perubahan itu akan meliputi tiga macam domain (ranah) yaitu pengetahuan (cognitive), sikap (affective), dan keterampilan (psychomotor).

Selanjutnya setiap ranah tersebut terbagi pula tas beberapa bagian atau tingkatan. Dalam makalah yang ringkas ini dicoba menguraikan dengan sederhana pembagian ranah pengetahuan (cognitive) sebagai berikut.

1. Knowledge (pengetahuan mengingat)

Termasuk ke dalam ranah ini kemampuan untuk mengenal fakta, terminologi, cara dan sebagainya.

2. Comprehension (pemahaman). Termasuk ke dalamnya menginterpretasikan, extrapolasi, dan penerjemahan maksud sesuatu.

3. Application (pemakaian/ penerapan). Ke dalam ranah ini tergolong kemampuan untuk menerapkan, mengembangkan, mengalihkan (tranfer).

4. Analysis (menganalisis). Diantaranya analisis tentang elemen, analisis hubungan, analisis prinsip- prinsip.

5. Synthesis (sistesis). Contoh dari tingkah laku untuk ranah ini seperti membentuk susunan (sturcture(s), membentuk pola (patren).

6. Evaluation (evaluasi). Termasuk ke dalam ranah ini pertimbangan (judgment) dengan menggunakan kriteria internal dan external evidence.

Kata Kerja Operasional sebagai indikator dari perubahan tingkah laku si belajar (change of behavior) dapat dilihat pada Tuckman, 1975, hal. 39-40).-

Dalam mempertimbangkan Content Validity (validitas isi) ke dua faktor yang telah diuraikan di atas yaitu materi kurikulum dan tingkatan ranah pengetahuan harus menjadi bahan pertimbangan. Supaya ke dua faktor tersebut terwakili ke dalam soal- soal yang akan ditanyakan sebagai alat ukur (instrument) dari kemampuan si belajar, maka pembuatan soal dirancang dengan menggunakan tabel yang disebut blue print (kisi- kisi penyusunan soal seperti contoh berikut ini.).

KISI-KISI PENYUSUNAN SOAL UJIAN
STATISTIK DASAR SEMESTER I 1990

| Ranah pengetahuan Materi | P e n g e t a h u a n (c o g n i t i v e) | | | | | | Jumlah f % |
|---|---|----------------|---------------|---------------|---------------|---------------|---------------|
| | Ingat- an | Pema- haman | Apli- kasi | Anali- sis | Sinte- sis | Evalu- asi | |
| Pengertian dan kegunaan Sta- tistik | 7 | 4 | 4 | - | - | - | 15 15 |
| Distribusi frekuensi | 4 | 3 | 3 | 3 | 2 | - | 15 15 |
| Macam- macam grafik | 3 | 4 | 5 | 4 | 2 | 2 | 20 20 |
| Nilai- nilai kecenderungan sentral | 3 | 4 | 4 | 5 | 2 | 2 | 20 20 |
| Ukuran varia- bilitas | 1 | 2 | 3 | 4 | 2 | 3 | 15 15 |
| Korelasi | 1 | 2 | 3 | 3 | 3 | 3 | 15 15 |
| Jumlah | - | - | - | - | - | - | 100 100 |

Perlu diingat bahwa ranah (domain) yang lebih tinggi otomatis

telah berisikan domain yang lebih rendah, akan tetapi sebaliknya ranah yang lebih rendah tidak akan mengandung domain (ranah) yang lebih tinggi.

Dengan menggunakan blue print (kisi- kisi) akan dapat diatur isi sel (petak- petak) pada tabel sehingga materi kurikulum yang telah disajikan dan tingkatan ranah pengetahuan (domain) dapat terwakili (representatif). Diasanya jumlah isi sel tersebut banyak ditentukan oleh pertimbangan para professional yang menguasai materi (bahan belajar yang bersangkutan atau expert judgment). Dengan cara tersebut di atas dapatlah dicapai content validity (validitas isi).

B. Construct Validity (validitas Konstruksi)

Yang dimaksud dengan konstruksi (construct) dalam makalah sama pengertiannya dengan " konsep ". Kalau kita berbicara tentang sesuatu objek atau maksud sebenarnya objek atau maksud tersebut berisikan konsep tentang objek atau maksud tersebut. Misalnya istilah beruang mengandung bermacam konsep. Pertama, mungkin beruang itu mengandung konsep mempunyai uang. Kemungkinan ke dua, beruang itu mempunyai konsep sejenis binatang yang ganas atau binatang buas. Ke -
mungkinan ke tiga istilah beruang itu mengandung konsep mempunyai ruang atau rongga. Dalam membuat soal yang mempunyai construct validity diharapkan pertanyaan (soal) tersebut mengandung konsep atau konstruksi yang sama dari setiap orang yang diuji atau sipembaca soal itu. Berikut ini diberikan sebuah contoh soal yang konstruksinya tidak valid untuk sebuah item test betul- salah.

B S Bunga ros tidak berduri.

Konsep bunga ros dalam soal tersebut tidak jelas .

Pertama, konsep bunga ros itu mungkin menunjukkan sebatang tumbuhan termasuk bangsa bunga-bungan ditanam manusia sebagai hiasan. ke dua, konsep bunga ros itu menunjukkan sekuntum bunga dari tanaman yang bernama bunga ros.

Dengan kata lain melihat konstruksi dari pertanyaan yang disajikan tidak menuju kepada suatu konsep atau construct yang sama bagi setiap pembaca. Contoh lain dari suatu soal yang tidak mempunyai validitas konstruksi (construct validity) adalah : Siapakah nama presiden Amerika Serikat ? Banyak nama president Amerika, yang mana dimaksudkan oleh si pembuat soal.

Untuk mendapatkan construct validity biasanya suatu soal direview oleh beberapa ahli bidang studi yang bersangkutan atau dilakukan try out (uji coba). Dengan kata lain kriteria untuk menentukan atau menetapkan construct validity biasanya didasarkan kepada pertimbangan para ahli (expert judgment).

C. Face Validity (validitas Pandangan)

Validitas pandangan ini ketepatan untuk menetapkan sesuatu bentuk seperti model arsitek, seni, bentuk telah yang cocok dengan ukuran. Tentu saja pandangan tersebut dilakukan oleh orang yang ahli di bidang yang bersangkutan. Contoh, dalam melakukan suatu kontes, perlombaan seni suara, seni musik, seni gambar, seni arsitek dan lain-lain yang serupa dengan itu banyak tergantung kepada pertimbangan (judgment) para ahli dalam bidang itu. Face validity ini lebih banyak mengandung unsur subjektif dibandingkan dengan validitas lainnya.

Menurut S. Arikunto Content validity dan construct validity disebut juga validitas logis (Arikunto 1990 hal. 64). Hal yang de-

demikian karena banyak pertimbangan- pertimbangan logis dari para ahli untuk menetapkan validitas yang disebutkan di atas. Menurut penulis makalah ketiga jenis validitas yang telah diuraikan diatas itulah yang lebih cocok dengan validitas logis yaitu: content, construct, serta face validity. Di lain pihak dikenal lagi validitas yang didasarkan kepada pengalaman yang diperoleh (empiris). Menurut S. Ari - kunto disebut juga validitas empiris (Akrikunto , 1990, hal. 64).

D. Predictive Validity (Validitas prediksi).

Suatu test dikatakan mempunyai ketepatan perkiraan (prediksi) bila hasil test tersebut dapat dipakai untuk memperkirakan kemampuan orang yang diuji dalam suatu academic performance tertentu di masa mendatang. Contoh, suatu test UMPTN (Ujian Masuk Perguruan Tinggi Negeri) dapat dikatakan mempunyai predictive validity apabila calon mahasiswa yang memperoleh hasil baik (tinggi) . memperoleh kesuksesan yang baik juga di lembaga tempat dia berkuliah. Demikian juga calon mahasiswa yang lulus pada UMPTN tetapi hasilnya hanya sekedar lulus tidak akan begitu sukses pada waktu kuliahnya nanti. Dengan kata lain bila dipantau hasil kemajuan mahasiswa di Perguruan tinggi misalnya dengan mempedomani IP (Indeks Prestasi) yang diperoleh mahasiswa pada akhir perkuliahannya dikorelasikan dengan nilai UMPTN pada waktu mulai akan masuk, bila terdapat korelasinya positif tinggi (besar dari .75) barulah dapat disebut bahwa UMPTN mempunyai predictive validity yang cukup memuaskan.

E. Concurrent Validity ("validitas dengan Ukuran yang telah ada sekarang")

Istilah concurrent berasal dari con atau co yang artinya "sama". Selanjutnya current artinya yang sedang ada sekarang . Dengan demikian concurrent validity berarti sama tepatnya dengan suatu ukuran

atau instrumen lain yang sekarang telah ada dipunyai. Contoh, ada suatu alat ukur yaitu EBANAS tahun yang lalu yang telah diketahui bahwa alat itu valid dan sekarang dipunyai oleh guru. Guru ingin mengetahui bahwa EBTA lokal yang disusun mereka di daerah apakah cukup valid? Cara untuk mengetahui validitasnya adalah dengan melakukan testing kepada murid yang akan ikut EBTA itu dengan dua jenis alat ukur yaitu EBANAS dan EBTA LOKAL. Hasil kedua test tersebut dikorelasikan. Apabila besarnya indeks korelasi antara kedua alat ukur tersebut positif tinggi (besar dari .75), dapatlah dikatakan bahwa EBTA LOKAL buatan guru itu valid (mempunyai concurrent validity dengan EBANAS). Selanjutnya dari penjelasan Tuckman di bawah ini akan lebih dapat difahami pengertian concurrent validity.

Concurrent validity tells whether the degree to which persons show evidence of a quality on a given test is reflected in or paralleled by their scores on another test of presumably the " same" characteristic. Often, for example, the concurrent validity of a test of mental ability is based on its relation to another tests of mental ability. (Tuckman , 1975, p.252).

Selanjutnya dapat dijelaskan bahwa predictive validity dan concurrent validity diperoleh berdasarkan pengalaman (empiris), tidak hanya didasarkan kepada logik semata, maka disebut juga validitas empiris (emperical validity).

III. RELIABILITY (KETETAPAN SUATU ALAT UKUR)

Dalam Bahasa Indonesia sering juga sering juga reliability itu ditukar istilahnya dengan reliabilitas. Suatu test dikatakan reliable dipercaya apabila hasil test tersebut menunjukkan angka yang konstan atau tetap setiap kali diteskan kepada testee yang sama apabila pengetahuan testee(s) tersebut tidak "ditambah". Istilah lain untuk reliable adalah keterandalan. Konsep reliabilitas tidak sama dengan

validitas. Reliability (reliabilitas) menunjukkan hasil yang tetap atau konstan dari suatu objek yang diukur; sebaliknya validitas berarti keteraturan memilih suatu alat serta cara menggunakan suatu alat ukur tersebut. Suatu alat ukur yang valid biasanya otomatis reliable, tetapi alat ukur yang reliabele belum tentu valid. Konsep validity telah dibahas pada uraian yang berlalu, selanjutnya dibawah ini dibahas teknik- teknik untuk mengetahui reliabilitas dari suatu alat ukur atau test.

A. Test - retest method

Teknik test retest ini dilaksanakan dengan test/ alat ukur yang sama dua kali pengukuran. Contoh, 14 Orang siswa diberikan test A pada waktu pertama hasilnya seperti tertera dalam kolom X, sedangkan pada test ke dua dengan alat ukur itu juga hasilnya seperti pada kolom Y. Dengan mengkorelasikan kedua hasil tersebut dapatlah diketahui indeks reliability dari test A.

Tabel 2

Korelasi antara test I dengan test II dengan alat yang sama.

| Subjek | Hasil I (X) | Hasil test II (Y) | X ² | Y ² | XY | |
|--------|-------------|-------------------|----------------|------------------------|------------------------|------------|
| 1 | 18 | 28 | 324 | 784 | 504 | |
| 2 | 18 | 30 | 324 | 900 | 540 | |
| 3 | 17 | 30 | 289 | 900 | 510 | |
| 4 | 17 | 26 | 289 | 676 | 442 | |
| 5 | 16 | 28 | 256 | 784 | 448 | |
| 6 | 16 | 24 | 256 | 576 | 384 | |
| 7 | 15 | 22 | 225 | 484 | 330 | |
| 8 | 15 | 20 | 225 | 400 | 300 | |
| 9 | 14 | 26 | 196 | 676 | 364 | |
| 10 | 14 | 22 | 196 | 484 | 308 | |
| 11 | 13 | 24 | 169 | 576 | 312 | |
| 12 | 13 | 18 | 169 | 324 | 234 | |
| 13 | 12 | 20 | 144 | 400 | 240 | |
| 14 | 12 | 18 | 144 | 324 | 216 | |
| N | 14 | EX= 210 | EY = 336 | EX ² = 3206 | EY ² = 8288 | EXY = 5132 |

Dengan menggunakan rumus

$$\begin{aligned}
 r &= \frac{EXY - \frac{(EX)(EY)}{N}}{\sqrt{\left[EX^2 - \frac{(EX)^2}{N}\right] \left[EY^2 - \frac{(EY)^2}{N}\right]}} \\
 &= \frac{5132 - \frac{(210)(336)}{14}}{\sqrt{\left[3206 - \frac{(210)^2}{14}\right] \left[8288 - \frac{(336)^2}{14}\right]}} \\
 &= \frac{92}{\sqrt{56 \times 224}} \\
 &= \frac{92}{112} = .82
 \end{aligned}$$

Jadi korelasi antara hasil test pertama dengan ke dua = .82 Selanjutnya arti umum dari indeks korelasi tersebut adalah sebagai berikut:

- Antara .80 sampai dengan 1.00 : sangat tinggi
- Antara .60 sampai dengan .79 : tinggi
- Antara .40 sampai dengan .59 : cukup
- Antara .20 sampai dengan .39 : rendah
- Antara 0.00 sampai dengan .19 : rendah sekali (disesuaikan dengan S. Arikunto 1990 hal. 71).

Untuk mengetahui signifikansi hubungan antara dua variable dengan indeks korelasi dapat dilihat dalam Tabel E yang terambil dari buku *Statistical Reasoning in Psychology and Education* oleh E. Minium 1978p. 539 -540 (lihat lampiran!).

Dengan memperhatikan hasil korelasi test retest, yaitu .82, dapat diartikan bahwa test A tersebut di atas adalah test reliabel.

B. Metode belah dua split- half method

Ada beberapa cara untuk membagi test tersebut menjadi dua bagian.

Sebelum dilakukan pembagian tentu jumlah item dalam test tersebut haruslah genap. Kalau sekiranya jumlah tersebut bukan merupakan bilangan genap terpaksa dibuang sebuah item.

Untuk membagi atas dua bagian cara yang sering dilaksanakan adalah membagi atas nomor genap dan ganjil, atau boleh juga teknik random yang lain. Berikut ini dicontohkan pembagian atas nomor genap dan ganjil.

Tabel 3

PEMBAGIAN HASIL TEST A ATAS NOMOR GENAP DAN GANJIL

| Subjek | N o m o r i t e m s | | | | | | | | | | | | | | | | | | | | Jumlah | |
|--------|---------------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|--------|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | GL | GP |
| A | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | 1 | 1 | 1 | - | - | 1 | 1 | - | - | - | 8 | 6 |
| B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 10 |
| C | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | 1 | 1 | 1 | - | 1 | - | - | 7 | 9 |
| D | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | 6 | 6 |
| E | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | 1 | 1 | 1 | 1 | 1 | 1 | - | 1 | - | 9 | 8 |
| F | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | 1 | 1 | - | 1 | - | - | 7 | 8 |
| G | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | - | - | 5 | 4 |
| H | 1 | 1 | - | - | 1 | 1 | 1 | 1 | - | - | 1 | 1 | - | 1 | 1 | - | - | 1 | - | - | 6 | 5 |
| I | 1 | - | - | - | 1 | - | 1 | 1 | 1 | 1 | 1 | 1 | - | 1 | - | - | - | 1 | - | 1 | 5 | 6 |
| J | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | 10 | 9 |

Dari Tabel 3 tersebut dapat dilihat bahwa siswa A mendapat score untuk semua items ganjil = 8, sebaliknya untuk semua items genap adalah 6; siswa B items ganjil = 10, dan item genap juga 10, demikian seterusnya.

Tabel 4

Korelasi items nomor genap dengan ganjil

| Siswa | items nomor ganjil (X) | items nomor genap (Y) | X^2 | Y^2 | XY |
|-------|--------------------------------|-------------------------------|--------------|--------------|----------|
| A | 8 | 6 | 64 | 36 | 48 |
| B | 10 | 10 | 100 | 100 | 100 |
| C | 7 | 9 | 49 | 81 | 63 |
| D | 6 | 6 | 36 | 36 | 36 |
| E | 9 | 8 | 81 | 64 | 72 |
| F | 7 | 8 | 49 | 64 | 56 |
| G | 5 | 4 | 25 | 16 | 20 |
| H | 6 | 5 | 36 | 25 | 30 |
| I | 5 | 6 | 25 | 30 | 30 |
| J | 10 | 9 | 100 | 81 | 90 |
| <hr/> | | | | | |
| N= 10 | EX= 73 | EY= 71 | $EX^2 = 565$ | $EY^2 = 539$ | EXY= 545 |

$$r = \frac{EXY - \frac{(EX)(EY)}{N}}{\sqrt{\left[EX^2 - \frac{(EX)^2}{N}\right] \left[EY^2 - \frac{(EY)^2}{N}\right]}}$$

$$r = \frac{545 - \frac{(73)(71)}{10}}{\sqrt{\left[565 - \frac{(73)^2}{10}\right] \left[539 - \frac{(71)^2}{10}\right]}} = \frac{26,7}{\sqrt{(53,1)(34,9)}}$$

$$r = \frac{26,7}{43,05} = .62 \text{ atau } 0,62.$$

Dari data tersebut berarti bahwa korelasi antara scores setengah be-
lahan test = .62. Untuk mengetahui indeks reliability dari test terse-
dipakai rumus Spearman Brown yaitu :

$$\text{Reliability of total test} = \frac{2 \times \text{reliability of half test}}{1 + \text{reliability of half test}}$$

(Tuckman 1975, p. 261)

$$\text{Jadi reliability dari test itu} = \frac{2 \times .62}{1 + .62} = \frac{1,24}{1,62} = .77 \text{ atau}$$

atau 0,77. Bila dilihat tabel E pada lampiran atau interpretasi hasil
korelasi ternyata merupakan korelasi yang cukup tinggi. Dengan kata
lain test tersebut cukup reliable.

C. Metode Bentuk Paralel (Equivalent Form)

Yang dimaksud dengan parallel equivalent forms ini adalah dua buah
test yang mempunyai tujuan, isi materi, dan tingkat kesukaran yang sa-
ma tetapi bentuknya berbeda. Kedua test tersebut diadministrasikan ke-
pada testee(s) yang sama. Untuk mengetahui reliabilitasnya maka cari-
lah atau hitunglah indeks korelasi ke dua buah test itu.

Berikut ini diberikan hasil (score) mentah test X yang equivalent
dengan test Y

| Subjek | test X | Test Y | X ² | Y ² | XY |
|--------|---------|---------|------------------------|------------------------|-----------|
| 1. | 18 | 28 | 324 | 784 | 504 |
| 2. | 18 | 30 | 324 | 900 | 540 |
| 3. | 17 | 30 | 289 | 900 | 510 |
| 4. | 17 | 26 | 289 | 676 | 442 |
| 5. | 16 | 28 | 256 | 784 | 448 |
| 6. | 16 | 24 | 256 | 576 | 384 |
| 7. | 15 | 22 | 225 | 484 | 330 |
| 8. | 15 | 20 | 225 | 400 | 300 |
| 9. | 14 | 26 | 196 | 676 | 364 |
| 10. | 14 | 22 | 196 | 484 | 308 |
| 11. | 13 | 24 | 169 | 576 | 312 |
| 12. | 13 | 18 | 169 | 324 | 234 |
| 13. | 12 | 20 | 144 | 400 | 240 |
| 14. | 12 | 18 | 144 | 324 | 216 |
| <hr/> | <hr/> | <hr/> | <hr/> | <hr/> | <hr/> |
| N= 14 | EX= 210 | EY= 336 | EX ² = 3206 | EY ² = 8288 | EXY= 5132 |

Hitung besarnya indek korelasi test X dengan Y . Indeks tersebut menunjukkan besarnya reliability dari test tersebut.

D. Reliability dengan RUMUS Kuder- Richardson Formula (K.R. 21)

Rumus tersebut adalah sebagai berikut

$$\text{Reliability} = 1 - \frac{\bar{X} (n - \bar{X})}{ns^2}$$

1 = bilangan konstan

\bar{X} = rata- rata score

n = Jumlah items

s = standard deviation -
a measure of variability
of a group of scores
around a mean .

(B. Tuckman 1975, p. 256)

Contoh pemakaian rumus , misalkan ada 12 items yang diberikan kepada 10 orang siswa yang scorenya adalah sebagai berikut:

| Siswa | Scores |
|----------|----------|
| A. | 12 |
| B. | 11 |
| C. | 10 |
| D. | 9 |
| E. | 9 |
| F. | 9 |
| G. | 7 |
| H | 6 |
| I | 5 |
| <u>J</u> | <u>2</u> |

$$N = 10 \quad \Sigma X = 80 \quad \bar{X} = 80/10 = 8, \quad s = 2,86 \quad (\text{dicari dengan calculator})$$

Bila dimasukkan ke dalam rumus K.R.21 maka dijumpai :

$$\begin{aligned} \text{Reliability} &= 1 - \frac{8(12-8)}{12(2,86)^2} \\ &= 1 - \frac{32}{9,84} = 1 - 0,33 \\ &= 0,67 \text{ atau } .67 \end{aligned}$$

Di atas telah diuraikan empat cara untuk mengetahui besarnya indeks reliability dari suatu test. Cara tersebut adalah yang biasanya dilakukan oleh orang , disamping itu tentu masih banyak cara lain. Di dalam makalah yang pendek ini penulis merasa cukup untuk diketahui penyusun soal bagaimana cara memperoleh tingkat reliability test yang mereka susun.

IV. USEBILITY (KEPRAKTISAN)

Suatu hal yang sangat perlu dipertimbangkan oleh penyusun soal

adalah usable yaitu daya guna atau praktisnya soal tersebut diperbuat atau dikerjakan oleh testee. Usability tersebut dapat ditinjau dari beberapa sudut pandangan.

A. Bahasa

Bahasa yang digunakan dalam menyusun soal sederhana dan mudah difahami (kecuali khusus untuk test tentang vocabulary). Susunan bahasa yang dipakai terutama sekali harus memperhatikan construct validity. Sedapat mungkin hindarilah kata-kata double negative (contoh tidak dalam keadaan tidak sakit).

Disamping itu penulisan tanda-tanda baca atau puntuasi harus tepat, demikian juga simbol-simbol rumus.

B. Format soal

Soal-soal disusun dalam format yang sedemikian rupa bentuknya sehingga memudahkan bagi testee membacanya dan tidak melewatkan/ menjemukan. Perhatikanlah contoh susunan nomor di bawah ini!

1

2

3

4

5

dst. Susunan seperti tersebut di

atas mungkin lebih baik dirubah menjadi berderet ke bawah sehingga tidak membosankan si pembaca soal, (lihat perbaikan di bawah)

1

2

3

4

5

6

dst.

C. Susunan Letak Soal

Soal- soal dalam satu unit tertentu sebaiknya diusahakan dalam satu halaman. Contoh, soal penjodohan, antara pernyataan dengan alternative pasangannya jangan sampai terpisah dalam satu halaman. Bila terpisah maka testee akan "membalik- balik"berikutnya sehingga meletihkan pemikiran; dan ini bukanlah tujuan test. Demikian juga dalam soal " Analisa sebab akibat " antara kalimat pernyataan dengan kalimat alasan usahakan dalam satu halaman yang tidak terpisah.

Semua faktor yang berhubungan dengan kepraktisan soal tersebut, apakah dijumpai dalam test tentang pengetahuan, sikap, atau praktek, dalam melakukan suatu keterampilan dikatakan usability dari test itu.

V. EKONOMIS DALAM PENYELENGGARAAN

Prinsip ekonomis ini adalah efisiensi dan efektifitas dari penyusunan dan pelaksanaan suatu test. Ada test yang bersifat individual dan ada pula yang dapat dilaksanakan secara klasikal. Hal ini bergantung kepada tujuan yang ingin dicapai. Sesuai dengan prinsip ekonomis untuk apa dilakukan test individual kalau dengan cara klasikal tujuan kita telah dapat tercapai. Dalam menyusun dan melaksanakan test; prinsip ekonomis yang perlu dipertimbangkan : pemakaian waktu, tenaga manusia, fasilitas(termasuk biaya) serta yang penting sekali adalah tujuan dari test. Pengertian melaksanakan suatu test tidak terbatas hanya kepada menjawab soal oleh testee, akan tetapi juga termasuk scoring, grading serta memberian informasi kepada testee sebagai feedback dari pekerjaan mereka.

VI. KESIMPULAN

Usaha seseorang pendidik atau pengajar tidak dapat dilepaskan dari evaluasi. Seorang pengajar tidak hanya menyusun program pengajaran serta menyampaikan program tersebut, tetapi juga bertanggung jawab untuk menilai hasil program yang diajarkannya. Sehubungan dengan itu seorang pengajar harus dapat menyusun berbagai bentuk instrument (alat ukur) yang tepat (valid), dan terpercaya (reliable) untuk suatu tujuan tertentu, disamping mempertimbangkan praktis dan ekonomisnya.

Sangat perlu diingat bahwa suatu test yang tepat untuk mengukur tujuan tertentu (misalnya untuk seleksi) belum tentu tepat (valid) untuk mengetahui kesulitan belajar si anak, (diagnostik).

Dibalik itu harus pula disadari oleh penyusun test, walaupun ia sudah berusaha menyusun test atau instrumen sesuai dengan yang diuraikan dalam makalah ini, ketahuilah bahwa sesungguhnya ketepatan (validity) yang mutlak berada di "tangan Tuhan"; sedangkan manusia diwajibkan hanya berusaha ke arah itu.

Seiring dengan itu, bahwa di dunia ini tidak ada sesuatu yang sesungguhnya konstan (reliable mutlak), sebab semua peristiwa yang terjadi di dunia ini hanya berlangsung satu kali. Satu-satunya yang reliable mutlak adalah " perkataan atau janji Tuhan! Semoga setiap penyusun test senantiasa mengingat/ mempedomani janji atau perkataan Tuhan itu.-

VII.

REFERENCES

- Arikunto. S, (1990), Dasar- Dasar Evaluasi Pendidikan, Jakarta, Bumi Aksara.
- Ary, Donald, et al, (1979), Introduction to Research in Education, Holt Rinehart and Winston, Inc. New York.
- Idris Zahara, Aliasar, dan Muhammad Arni, (1979), Pengukuran dan Penilaian Pendidikan , FIP IKIP Padang.
- Lindvall , and Nitko,(1975), Measuring Pupil Achievement and Aptitude, Harcourt Brace Jovanovich, Inc., New York.
- Minium W. Edward, (1978), Statistical Reasoning in Psychology and Education, John Willey & Sons. New York.
- Tuckman W.B,(1975), Measuring Educational Outcomes, Fundamental of Testing , Harcourt Brace Jovanovich, New York.