

TEKNIK MEMBUAT SOAL

Oleh:

Dr. Syafri Anwar, M.Pd
(Universitas Negeri Padang-UNP)

16-1-2009
HR
K1
29 HR / 2005 - t(1)
3716 2711 Anw t.1
education - Tests and Measurements

Makalah Disampaikan pada Kegiatan Pelatihan Instruktur Sentra
Pendidikan BRI Selindo
Di Padang, Tanggal: 26 s/d 28 Juni 2007

TEKNIK MEMBUAT dan ANALIS SOAL¹

Oleh: Syafri Anwar²

Abstrak

Tulian ini bertujuan menjelaskan tentang bagaimana teknik membuat soal yang khususnya soal objektif bentuk pilihan ganda. Selama ini ada kecenderungan pembuat tes memandang bahwa membuat soal itu mudah, apalagi kalau soal itu adalah soal objektif bentuk pilihan ganda. Namun pada kenyataannya banyak setelah soal direspon oleh peserta tes, banyak masyarakat (orang tua, guru, dan siswa sendiri) yang mempersoalkan rendahnya mutu soal yang dibuat. Ini berarti bahwa membuat soal ternyata tidak mudah, karena untuk menghasilkan satu butir soal yang bermutu diperlukan beberapa persyaratan penting. Di antara persyaratan itu, ada tiga syarat mutlak yang mesti dipenuhi yaitu: (1) syarat materi, (2) syarat konstruksi soal, dan (3) syarat bahasa. Selain itu persyaratan tersebut ada lagi beberapa persyaratan lain yang ikut mendukung suatu mutu soal. Persyaratan yang dimaksud adalah persyaratan administratif, dan persyaratan sistem penyelenggaraan ujian itu sendiri dan mental orang-orang berkepentingan. Artinya, bagaimanapun baiknya butir soal dirancang, tetapi apabila dalam penyelenggaraannya tidak benar, ada kecurangan, maka soal yang baik itu sulit diprediksi. Untuk itu, kalau soal yang dibutuhkan benar-benar objektif, maka selain soal itu dirancang mengikuti kaidah-kaidah yang benar, dalam proses pengerjaan dan penilaiannya juga dilakukan cara-cara yang benar pula.

Kata kunci: Tes, soal objektif pilihan ganda, syarat mutu soal

Pendahuluan

Guru mempunyai tugas pekerjaan yang cukup kompleks, khususnya dalam proses belajar mengajar di sekolah. Tugas-tugas ideal guru antara lain; membuat perencanaan pembelajaran, menyajikan pelajaran kepada peserta

¹ Makalah disampaikan pada kegiatan pelatihan instruktur widyaswara BRI tingkat nasional, di Sendik BRI Padang, 2007.

² Dosen jurusan Pendidikan Geografi FIS UNP Padang

didik, memilih metode yang tepat, dan terakhir adalah mengevaluasi hasil belajar. Dari serangkaian tugas guru itu, mengevaluasi / mengujikan soal adalah tugas yang membutuhkan kiat, keterampilan, dan keahlian khusus. Secara sepintas lalu membuat soal kelihatannya mudah, bahkan ada anggapan kalau membuat soal itu dapat dikerjakan pada saat ujian akan dimulai. Anggapan demikian tentunya boleh-boleh saja, asalkan soal yang dikembangkan secara sembrono itu tidak digunakan untuk kepentingan pengambilan kebijakan seperti untuk kenaikan kelas, lulus sekolah, atau memasuki suatu pekerjaan. Yang ingin dijelaskan di sini adalah, untuk mendapatkan soal yang baik diperlukan kiat atau teknik-teknik tertentu sebagaimana yang akan dipaparkan pada bagian selanjutnya di dalam makalah ini.

Dari berbagai kasus pelaksanaan ujian, kita sering mendengar adanya keluhan baik dari guru maupun dari peserta tes setelah mereka membaca dan merespon soal-soal yang diberikan. Contohnya, Asosiasi Guru Matematika Indonesia di Bandung (Kompas, 27 April 2007) yang menyatakan bahwa mutu soal ujian nasional tahun 2007 merosot (Kompas, April 2007: <http://www2.kompas.com/>). Kasus lain misalnya, pernyataan seorang siswi SMA 68 Jakarta menanggapi mutu soal geografi usai menghadapi Ujian Nasional (UN), "gilaa, susah banget, benar-benar *ngak nyangka*, soal yang keluar *kaya* gitu, *nggak* familiar banget, jauh dari yang selama ini kita latihan," (Kompas, 23 April 2008: <http://www.kompas.com/>). Beberapa pernyataan lain dari siswa dengan sumber yang sama antara lain, "soalnya nggak sesuai dengan SKL", "kecewa banget liat soalnya", dan berbagai keluhan serupa dari sumber-sumber lain.

Permasalahan seperti yang dikemukakan di atas kiranya sudah cukup sebagai alasan untuk mengemukakan bahwa membuat soal ternyata tidak semudah yang dibayangkan oleh banyak orang. Case dan Swanson (2002) mengidentifikasi banyak hal sebagai penyebab rendahnya mutu soal di antaranya; pernyataan soal yang terlalu panjang, bahasa soal mengaburkan pemahaman siswa, dan ada bias dalam jawaban soal karena perbedaan kata-kata. Oleh sebab itu, terutama orang-orang yang punya kepedulian dalam bidang kependidikan seperti guru, widiyaswara, pembuat soal pada tingkat/lembaga/instansi tertentu, perlu mengetahui bagaimana teknik membuat soal yang bermutu. Pengetahuan ini diharapkan dapat membantu pihak-pihak yang berkepentingan dalam menyusun soal, baik di sekolah maupun di lembaga non-sekolah pada masa yang akan datang.

Pembahasan

Pada hakikatnya ada dua kegiatan yang dilakukan untuk mengevaluasi tingkat keberhasilan dalam penguasaan tertentu yaitu tes dan non test. Dalam kegiatan pendidikan dan pengajaran di sekolah, secara umum tes dan non tes dilaksanakan untuk mendapatkan informasi tentang karakteristik dan kinerja siswa (Wiersma & Jurs, 1990). Dalam pelaksanaannya antara tes dan nontes berbeda, di mana tes dilaksanakan dalam bentuk tes tulis biasa dan lebih dikenal *paper and pencil test*, sedangkan nontes dilaksanakan dalam bentuk wawancara, observasi, dan telli tingkah laku (*behavioral tellies*). Tes yang dilaksanakan dalam bentuk *paper and pencil test* pada dasarnya berhubungan dengan pembuatan soal, apakah itu soal objektif maupun soal esay. Soal objektif adalah soal yang terdiri dari satu pernyataan soal dan beberapa alternatif / pilihan jawaban (Mc Beath, 1992). Soal esai adalah soal yang

menghendaki jawaban terbatas (*restricted response*) maupun jawaban luas (*extended response*) yang cenderung bertujuan untuk mendapatkan informasi tentang kemampuan tingkat tinggi siswa (Popham, 1995).

Wiersma & Jurs (1990) mengemukakan tujuh fungsi tes yaitu: (1) fungsi prediksi, (2) menilai perbedaan individu, (3) pengambilan keputusan yang objektif, (4) estimasi domain, (5) mengetahui tingkat penguasaan, (6) diagnosa, (7) menilai perkembangan belajar melalui penilaian sebelum-sesudah (*before-after*). Berfungsi prediksi bertujuan mengetahui sejauhmana tingkat keberhasilan peserta didik pada masa yang akan datang dalam mewujudkan suatu prestasi. Tes yang berfungsi prediktif ini misalnya Tes Bakat Scholastik (TBS) atau *Scholastic Aptitude Test (SAT)*. Tes berfungsi menilai perbedaan individu bertujuan mengetahui seberapa besar perbandingan antara siswa yang menguasai informasi instruksional dengan yang belum/tidak menguasai sama sekali. Tes berfungsi pengambilan keputusan terutama diperlukan oleh pembuat tes untuk menentukan apakah seseorang dapat melanjutkan penguasaannya pada tahap yang lebih tinggi atau tidak, misalnya untuk naik kelas dan kelulusan sekolah. Demikian juga halnya dengan estimasi domain, diagnosis, yang muaranya adalah menilai perkembangan belajar siswa, lalu atas atas penilaian itu ditetapkan kebijakan-kebijakan untuk mencapai target yang diharapkan. Case & Swanson (2002) mengemukakan tujuan test yang lain yaitu: (1) mengkomunikasikan material-material penting dalam pembelajaran, (2) mengidentifikasi kelemahan siswa untuk tujuan remedial, (3) mengidentifikasi pencapaian akhir untuk kepentingan kebijakan, (4) mengidentifikasi keemahan kurikulum. Selain itu Djaali (2002) mengemukakan bahwa fungsi tes selain alat untuk mengukur prestasi tetapi

juga sebagai motivator kepada peserta didik dan perbaikan kualitas pembelajaran.

Tulisan ini diarahkan untuk membahas tentang soal objektif, khusus dalam bentuk pilihan ganda (*multiple choice*). Sebagaimana dikemukakan di atas, secara sederhana tes objektif adalah tes yang terdiri dari pernyataan soal dan sejumlah alternatif jawaban, yang menghendaki satu jawaban benar di antara beberapa alternatif jawaban yang ditawarkan itu. Sepintas lalu soal objektif pilihan ganda kelihatan sederhana, dan mungkin dianggap tidak terlalu sukar untuk membuatnya. Namun perlu diingat bahwa membuat soal seperti ini membutuhkan kiat dan teknik tertentu, sehingga diperoleh soal objektif yang bermutu.

Sebelum masuk kepada pembuatan soal, seorang evaluator hendaknya mengetahui beberapa hal tentang aspek-aspek evaluasi secara umum. Salend. Thorndike, dan Hagen dalam Ratt (1986) mengemukakan ada enam aspek yang dimaksudkan yaitu; aspek informasi tentang tes, alat bantu, pertimbangan peserta tes, reliabilitas dan validitas, administrasi dan penyekoran, dan skala/norma. Masing-masing aspek memiliki beberapa rincian, sebagaimana terdapat pada Tabel -1 berikut.

Tabel 1: Pedoman untuk tes evaluasi

No	Aspek-aspek evaluasi yang perlu dipertimbangkan dalam merancang soal	Rincian
1	Informasi tentang tes	a. Nama tes b. Pembuat tes c. Tanggal tes d. Lama pengerjaan soal
2	Alat Bantu	a. Apakah alat Bantu sudah cukup jelas (tujuan dan kualifikasinya)? b. Apakah alat bantu menuntun ke arah interpretasi siswa yang benar c. Apakah pernyataan soal berhubungan dengan apa yang dilihat
3	Pertimbangan peserta tes	a. Bagaimana kemampuan awal peserta tes b. Dalam bahasa apa tes dirancang c. Apakah tingkat bahasa cocok dengan kemampuan peserta tes d. Bagaimana item tes dirancang e. Apakah peserta tes dapat beradaptasi dengan soal f. Apakah tes sudah bebas dari bias deskriminasi (etnis dan seks) Apakah tes cocok untuk individu atau kelompok
4	Reliabilitas dan validitas	a. Bagaimana validitas tes dan bagiannya. b. Bagaimana daya beda soal c. Bagaimana tingkat kesukaran soal d. Bagaimana keberfungsian pengecohnya
5	Administrasi dan penyekoran	a. Apakah petunjuk tes sudah jelas

		b. Apakah penyekorannya sudah disiapkan	teknik sudah
6	Skala dan norma	a. Skala apa yang digunakan b. Apakah skala itu tepat atau tidak c. Apakah norma yang dipakai sudah sesuai dengan populasi dan tujuan tes d. Apakah soal sudah didiskusikan	

Sumber : Salend Dkk dalam Ratt (1986)

Selain kriteria yang dikemukakan oleh Ratt di atas, secara lebih rinci Pusat Pengujian Nasional (Pusjian) sudah mengeluarkan pedoman atau ketentuan membuat soal, khususnya soal bentuk pilihan ganda. Lebih jelas ketentuan ini dapat dilihat melalui Tabel-2 berikut.

Tabel 2 : Ketentuan dalam membuat soal objektif bentuk pilihan ganda

BIDANG PENELAAHANN	KRITERIA PENELAAHAN EAVLUASI	KETERANGAN
A Materi	<ol style="list-style-type: none"> 1. Soal sesuai dengan indikator 2. Pilihan jawaban hogen dan logis 3. Hanya ada satu kunci jawaban yang tepat 	
B Konstruksi soal	<ol style="list-style-type: none"> 1. Pokok soal dirumuskan dengan singkat, jelas, dan tegas 2. Rumusan pokok soal dan pilihan merupakan pernyataan yang diperlukan 3. Pokok soal tidak memberi petunjuk ke kunci soal 4. Pokok soal sebaiknya menghindari pernyataan negatif ganda 5. Pilihan jawaban homogen dan 	Semakin sering guru membuat tes, kriteria ini semakin melekat dalam

	<p>logis ditinjau dari segi materi</p> <ol style="list-style-type: none"> 6. Panjang pilihan jawaban relatif sama 7. Hindari memberikan alternatif jawaban, "semua jawaban di atas salah, aau semua jawaban di atas benar". 8. Opsi jawaban yang berbentuk angka; usahakan diurutkan secara kronlogis (besar-kecil atau kecil besar 9. Gambar grafik, tabel, diagram, dan sejenisnya dibuat sejelas mungkin dan berfungsi 10. Butir soal tidak bergantung/ ditentukan oleh jawaban sebelumnya 	ingatannya, dan akhirnya akan menjadi suatu kebiasaan saja
C Bahasa	<ol style="list-style-type: none"> a. Menggunakan kaidah bahasa yang baik/ EYD b. Tidak menggunakan bahasa setempat c. Tidak mengulang kata yang sama 	

Sumber : Pusjian, Depdiknas (2000/2001)

Selanjutnya McBeath (1992) mengemukakan ada tujuh ketentuan dalam membuat soal objektif bentuk pilihan ganda. Ketentuan McBeath ini lebih dikenal dengan istilah *seven roles for items construction*. Ketentuan tersebut antara lain: (1) pernyataan soal hanya terdiri dari satu *statement* saja, (2) hindari pernyataan yang tidak berhubungan dengan materi, (3) hanya ada satu jawaban benar di antara alternative jawaban, (4) hindari pernyataan negative, (5) pernyataan soal tidak memberi petunjuk kepada jawaban benar, (6) hindarkan pilihan jawaban, "semua jawaban di atas benar atau semua jawaban di atas salah", dan (7) panjang pengecoh relative sama.

Selanjutnya penulis mencoba mengambil beberapa contoh atau kasus soal yang dianggap keluar dari ketentuan teknik penyusunan soal

sebagaimana dikemukakan beberapa pakar tersebut di atas. Lebih rinci dapat dilihat melalui lampiran-1 makalah ini.

Dari beberapa uraian tersebut di atas, diperoleh gambaran bahwa untuk membuat soal bermutu ada beberapa tahapan yang mesti dilakukan oleh perancang soal. Tahapan yang dimaksud antara lain: tahap perencanaan, tahap membuat tes, tahap pelaksanaan tes, tahap evaluasi mutu soal. Tahap perencanaan dilaksanakan sebelum pembuatan tes. Kegiatannya antara lain: mengevaluasi keterlaksanaan kegiatan pembelajaran, mengidentifikasi materi, membuat kisi-kisi soal berdasarkan Kompetensi Dasar (KD) dan Standar Kompetensi (SK) yang sudah dirancang. Tahap membuat tes kegiatannya antara lain menulis soal itu sendiri, menilai validitas isi bersama pakar (*expert judgment*), dan memperbaiki soal. Dalam pelaksanaan tes, kegiatan penting adalah melakukan pengawasan sehingga hasil tes benar-benar murni. Pada bagian akhir adalah menilai mutu soal. Penilaian dapat dilakukan menggunakan analisis statistik sederhana atau statistik deskriptif prosentase. Penilaian mutu soal ditujukan untuk mengetahui taraf kesukaran soal (lampiran-2), indeks daya beda (lampiran-3), keberfungsian pengecoh (lampiran-4), dan validitas soal secara empirik (lampiran-5). Penilaian terhadap validitas empiris sudah cukup untuk membuktikan bahwa suatu butir soal sudah memiliki taraf kesukaran yang baik, memiliki indeks daya beda, dan keberfungsian pengecoh yang baik.

Penutup

Berdasarkan pembahasan di atas dapat disimpulkan bahwa, ada tiga hal pokok yang harus diperhatikan dalam merancang soal khususnya soal bentuk pilihan ganda. Ketiga hal pokok yang dimaksud antara lain: (1) syarat materi,

(2) syarat konstruksi soal, dan (3) syarat bahasa. Selain itu persyaratan tersebut ada lagi beberapa persyaratan lain yang ikut mendukung suatu mutu soal. Persyaratan yang dimaksud adalah persyaratan administratif, dan persyaratan sistem penyelenggaraan ujian itu sendiri serta sikap mental orang-orang berkepentingan.

Disarankan kepada guru dan dosen atau perancang soal lainnya, agar dalam setiap penyusunan soal perlu ada sikap kritis terhadap soal yang telah dibuat, sehingga dapat direvisi sebelum soal itu dijawab oleh siswa.

Selanjutnya, apabila siperancang soal akan membuat soal, hendaklah sipembuat soal merujuk kepada pedoman-pedoman penyusunan soal yang sudah dikembangkan oleh instansi terkait, terutama yang berada di bawah naungan Departemen Pendidikan Nasional (Depdiknas).

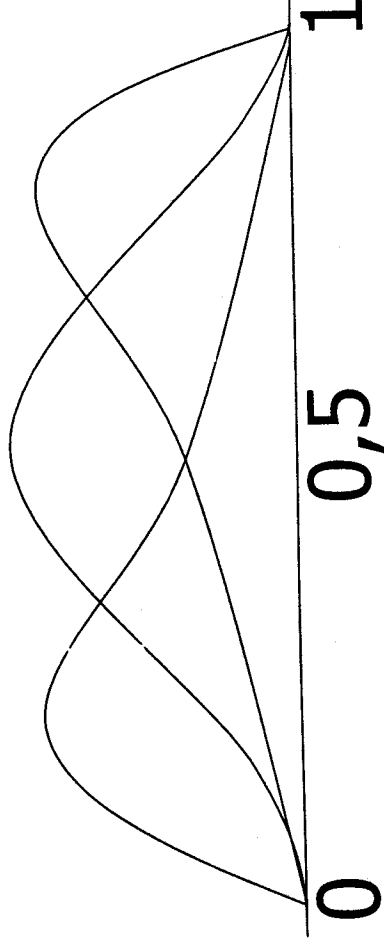
Daftar Rujukan

- Case, Susan.M dan Swanson, David.B. 2002. *Constructing Written Test Questions for the Basic and Clinical Science*, 3rd edition. (<http://cte.umdnj.edu/>). Philadelphia: National Board of Medical Examiners.
- Depdiknas.2000. *Bahan Penataran Pengujian Pendidikan*. Jakarta: Pusat Pengujian.
- Depdiknas. 2000. *Kebijakan Umum Program Pengembangan Pendidikan Menengah Umum (Dikmenum)*. Jakarta: Direktorat Dikmenum.
- McBeath, RJ. 1992. *Instruction and Evaluating in Higher Education. A Guide book for Planning Learning Outcomes*. New Jersey: Englewood Cliffs.
- Popham, WJ. 1995. *Classroom Assessment, What Teachers Need to Know*. Boston: Allyn & Bacon.
- Wiersma.W & Jurs, SG. 1990. *Educational Measurement and Testing*. Boston: Allyn & Bacon.

Lampiran 2

DERAJAT KESUKARAN ITEM

$$P = 0 - 1$$



$$P = \frac{B}{N} = \frac{6}{8} = 0,75$$

lampiran 3

DAYA PEMBEDA ITEM

$$D = P_A - P_B$$

$$P_A = \frac{B_A}{J_A}$$

$$P_B = \frac{B_B}{J_B}$$

$< 0,20 = jelek$

$0,20 - 0,40 = cukup / sedang$

$0,41 - 0,70 = baik$

$0,71 - 1,00 = sangat baik$

Minus = sangat jelek

Keterangan:

- *PA = Proporsi kel.atas yang menjawab benar*
- *PB = Proporsi kel.bawah menjawab benar*
- *JA = Jumlah kel.atas*
- *BA = Banyak kel.atas yang menjawab benar*
- *BB = Banyak kel.bawah menjawab benar*
- *JB = Jumlah kel.bawah*

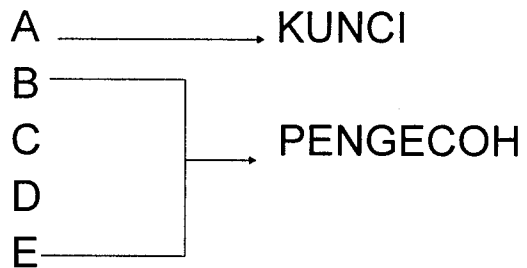
371.271
SYA
T-1

Agil.Ha/2005 = t (1)

371.271
Anw
t-1

Lampiran 4
ANALISIS FUNGSI DISTRAKTOR

- Pertanyaan :
- Alternatif jawaban:



Ketentuan:

Min. 5% peserta terkecoh pada masing- masing distraktor
Ex. 50 teste dengan 40 soal

ITEM	ALTERNATIF JAWABAN					KET.
	A	B	C	D	E	
1	4	6	5	30	5	
2	1	44	2	1	2	
3	1	1	10	1	37	

Aplikasi

- Butir 1 = A = $4/50 \times 100 = 8\%$
- B = $6/50 \times 100 = 12\%$
- C = 10%
- D = ?
- Butir 2 = A = ?
- dst
- Butir 3 = dst



Lampiran 5

Rumus product moment

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\{N\sum x^2 - (\sum x)^2\}\{N\sum y^2 - (\sum y)^2\}}}$$

- Tentukan validitas soal untuk item no: 6
- Buat tabel pembantu

Tabel Pembantu untuk validitas item no: 6

x	y	x^2	y^2	xy		
1	8	1	64	8		
0	5	0	25	0		
1	4	1	16	4		
1	5	1	25	5		
1	6	1	36	6		
0	4	0	16	0		
1	7	1	49	7		
1	8	1	64	8		
$\sum x=6$	$\sum y=47$	$\sum x^2=6$	$\sum y^2=295$	$\sum xy=38$	$(\sum x)^2=36$	$(\sum y)^2=2209$

Aplikasikan ke rumus

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\{N\sum x^2 - (\sum x)^2\}\{N\sum y^2 - (\sum y)^2\}}}$$

$$r_{xy} = \frac{8.38 - 6.47}{\sqrt{\{8.6 - 36\}\{8.295 - 2209\}}}$$

$$r_{xy} = 0.52$$

2. Rumus point biserial (r_{pbis})

$$r_{pbis} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

r_{pbis} = koefisien korelasi biserial

M_p = rerata skorsubyek yang menjawab benar item

M_t = rerata skor total

SD_t = S tan dar deviasi skor total

p = proporsi siswa yang menjawab butir dengan benar

q = proporsi siswa yang salah ($q = p - 1$)

Aplikasi ke rumus

$$r_{pbis} = \frac{Mp - Mt}{SD_t} \sqrt{\frac{p}{q}}$$

$$Mp = \frac{8+4+5+6+7+8}{6} = 6.33$$

$$Mt = \frac{8+5+4+5+6+4+7+8}{8} = 5.88$$

$$SD_t = \sqrt{\frac{\sum x_t^2}{N} - \left(\frac{\sum x_t}{N}\right)^2}$$

$$SD_t = \sqrt{\frac{295}{8} - \left(\frac{47}{8}\right)^2}$$

$$SD_t = 1.54$$

	$r_{pbis} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{P}{q}}$
	$r_{pbis} = \frac{6.33 - 5.88}{1.54} \sqrt{\frac{0.75}{0.25}}$
	$r_{pbis} = 0.51$
	$r_{xy} = 0.52$

	<p>2. Reliabilitas tes (ketetapan, kekonsistensian)</p> <p>a. Reliabilitas tes esay (<i>rumus alpha</i>)</p> <p>b. Reliabilitas tes obyektif</p>
	<p>* Single test-single trial method (rumus: spearman – Brown, Flanagan, Rulon, kuder Richardson, dan C.Hoyt).</p> <p>* Test-retest (tes ulang, 2 X dalam kesempatan berbeda (korelasi rank-order Spearman)).</p> <p>* Alternatif form (double test-double trial/ paralel) menggunakan rumus <i>product moment pearson</i> atau korelasi rank order Spearman (untuk N < 30)</p> <p style="padding-left: 40px;">$r_{ii} \geq 0,70 = \text{reliabel}$</p> <p style="padding-left: 40px;">$r_{ii} < 0,7 = \text{unreliabel}$</p>

JALAN PINTAS

MENGGUNAKAN PROG:

1. ANATES
2. Aplikasi rumus ke dalam Excel, dan
3. Iteman (*item analysis*)



Aplikasi R_{pbis} ke Excel

- Pi: $=SUM(B?:B?)/?$
- Qi: $=1-B?$
- Rerata (Xi):
 $=SUMIF(B?:B?;">0";$L$?:$L$?)/SUMIF(B?:B?;">0")$
- Rerata (Xt): $=AVERAGE($L$?:$L$?)$
- SD: $=STDEV($L$?:$L$?)$
- R_{pbis} : $=((B?-B?)/B?)*(SQRT(B?/B?))$
- *r-kritis:* *lihat pada tabel r-kriteria*
- Status:
 $=IF(B?>B?;"Valid";IF(B?<B?;"Drop"))$