

ANALISIS BUTIR DAN KONSISTENSI INTERNAL TES *)

Makalah

Oleh
I Wayan Santyasa **)

*) Disajikan dalam work shop bagi para Pengawas
dan Kepala Sekolah Dasar di Kabupaten Tabanan
Pada Tanggal 20-25 Oktober 2005
di Kediri Tabanan Bali

***) Adalah Dosen Tetap Jurusan Pendidikan Fisika
Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam
Institut Keguruan dan Ilmu Pendidikan Negeri
Singaraja

ANALISIS BUTIR DAN KONSISTENSI INTERNAL TES

I Wayan Santyasa

Jurusan Pendidikan Fisika IKIP Negeri Singaraja

Apabila hasil-hasil tes buatan Guru digunakan untuk tujuan-tujuan pembuatan keputusan, misalnya untuk *seleksi*, *klasifikasi*, *sertifikasi*, dan *pengakuan*, maka tes itu hendaknya mampu memberikan informasi seakurat mungkin. Untuk maksud tersebut, analisis tes menjadi sangat penting bagi para Guru dalam praktik sehari-hari. Hal ini sering tidak mampu dilakukan, mungkin disebabkan oleh beberapa alasan: (1) Guru tidak selalu memahami atau mengabaikan arti pentingnya evaluasi yang tepat, (2) Guru tidak siap mengenai metode untuk menganalisis tes, (3) Guru mungkin merasakan bahwa analisis tes tersebut terlalu banyak menghabiskan waktu. Jika ini terjadi, yang berarti bahwa konsep keefektifan tes tidak menjadi perhatian para Guru, maka keputusan yang ditetapkan berdasarkan tes yang dibuatnya boleh jadi tidak mampu menggambarkan keputusan yang benar.

Untuk tes prestasi belajar (*achievement test*), analisis tes yang dimaksud, paling tidak menyangkut validitas isi (*content validity*), *analisis butir*, dan konsistensi internal (*internal consistency*) baik butir maupun tes.

Validitas Isi

Gay (1987:129) menyatakan bahwa validitas isi (*content validity*) adalah derajat pengukuran yang mencerminkan domain isi yang diharapkan. Validitas isi penting untuk tes hasil belajar (*achievement test*). Suatu skor kurang bahkan tidak mencerminkan hasil belajar siswa apabila instrumen tidak mampu mengukur secara komprehensif apa yang telah dipelajari oleh siswa.

Prosedur yang hendak ditempuh agar suatu tes hasil belajar mampu mencerminkan domain isi secara komprehensif adalah dengan menyusun kisi-kisi tes. Tabel 1 menyajikan contoh kisi-kisi tes.

Tabel 1 Contoh kisi-kisi tes

Pokok Bahasan/Sub Pokok Bahasan	No Butir Perjenjang Kemampuan						Jml Butir
	C1	C2	C3	C4	C5	C6	
1.....							
1.1	1	2, 3	4, 5	6	7	8	8
1.2	9, 10	11, 12	13	14	15	16	8
1.3	17	18, 19	20	21	22	23	8
.....
n							
Jumlah Butir	5	25	5	5	5	5	50

Keterangan: C1 = hafalan, C2 = pemahaman, C3 = penerapan, C4 = analisis, C5 = sintesis, C6 = evaluasi

Validitas isi suatu tes hasil belajar tidak terlalu penting untuk dikuantifikasi. Validitas isi cukup diestimasi berdasarkan pertimbangan ahli isi. Sebagai ahli isi dapat ditunjuk seorang guru pada bidang studi yang sama yang memiliki kualifikasi dan pengalaman kerja yang cukup. Pertimbangan ahli tersebut dianggap cukup representatif sebagai dasar untuk memutuskan bahwa tes yang dikembangkan telah memenuhi syarat validitas isi. Di samping pemeriksaan oleh teman sejawat yang dianggap sebagai ahli, tes juga perlu diuji keterbacaannya ditinjau dari pemakai (siswa). Prosedur ini dilakukan melalui uji kelompok kecil dan kelas yang sesungguhnya.

Analisis Butir Tes Pilihan Ganda

Analisis butir adalah proses menguji respon-respon siswa untuk masing-masing butir tes dalam upaya menjustifikasi kualitas item. Kualitas item, khususnya direpresentasi oleh *daya beda item*, *tingkat kesukaran item*, dan khusus untuk tes pilihan ganda tidak kalah pentingnya adalah *keefektifan pengecoh* (Mehrens & Lehmann, 1984).

Untuk melakukan analisis butir, Guru dianjurkan melakukan langkah-langkah berikut:

1. Periksa hasil pekerjaan siswa dan berikan skor secara teliti dan cermat, kemudian masukkan skor-skor tersebut dalam Tabel 2.

Tabel 2 Skor butir tes pilihan ganda

No. Resp	Skor Perbutir										Skor Total
	1	2	3	4	5	6	7	8	9	10	
1	0	1	1	0	0	0	1	1	0	0	4
2	1	1	1	1	1	1	1	1	0	0	8
3	0	1	0	1	1	1	1	0	1	1	7
4	0	0	0	1	1	1	0	1	1	1	6
5	1	1	1	1	1	0	0	0	0	0	5
6	1	1	1	0	0	0	0	0	0	0	3
7	1	1	1	1	1	0	0	0	0	0	5
8	0	0	1	1	1	0	0	0	1	1	5
9	1	1	1	0	0	0	0	0	0	0	3
.
n

2. Urutkan skor siswa dari yang tertinggi hingga yang terendah. Langkah ini dilakukan hanya mengubah skor-skor pada Tabel 2 dengan mengurutkan dari skor tertinggi hingga terendah, seperti pada Tabel 3.

Tabel 3 Skor butir tes pilihan ganda setelah skor responden diurutkan

No. Resp	Skor Perbutir										Skor Total
	1	2	3	4	5	6	7	8	9	10	
2	1	1	1	1	1	1	1	1	0	0	8
3	0	1	0	1	1	1	1	0	1	1	7
4	0	0	0	1	1	1	0	1	1	1	6
5	1	1	1	1	1	0	0	0	0	0	5
7	1	1	1	1	1	0	0	0	0	0	5
8	0	0	1	1	1	0	0	0	1	1	5
1	0	1	1	0	0	0	1	1	0	0	4
6	1	1	1	0	0	0	0	0	0	0	3
9	1	1	1	0	0	0	0	0	0	0	3
.
n

3. Tetapkan kelompok atas (KA) dan kelompok (KB) dari skor-skor siswa yang telah diurutkan seperti pada Tabel 2. Jumlah KA atau KB disesuaikan dengan jumlah responden seluruhnya. Untuk jumlah responden relatif banyak (sekitar 100), dapat digunakan angka 30%, 27%, 25%. Tetapi untuk jumlah responden relatif sedikit, jumlah tersebut dapat disesuaikan, bahkan jika hanya 40 orang, maka KA atau KB dapat ditetapkan 20.
4. Hitung jumlah siswa baik pada KA maupun pada KB untuk masing-masing pilihan jawaban.
5. Sebagai contoh, misalkan jumlah responden seluruhnya adalah 40, maka KA = 20 dan KB = 20. Jumlah siswa pada KA dan KB yang menjawab pada masing-masing pilihan disajikan pada Tabel 4.

Tabel 4 Sebaran jumlah siswa pada masing-masing pilihan

Pilihan	A	B	C*	D	E	tidak memilih
KA	0	4	12	1	3	0
KB	4	2	8	5	1	0

*) berarti kunci jawaban

6. Hitung Ideks Kesukaran Butir (IKB) dengan formula

$$IKB = \frac{R}{T} \times 100\% \quad (1)$$

dengan IKB = Ideks Kesukaran Item, R = jumlah responden yang menjawab benar, dan T = jumlah responden seluruhnya. IKB dapat bernilai 0,00-1,00; IKB: 0,00 – 0,20 adalah *sangat sukar*, 0,20-0,40 *sukar*, 0,40-0,60 *sedang*, 0,60-0,80 *mudah*, dan

0,80-1,00 *sangat mudah*. Biasanya butir yang ditoleransi sebagai tes standar adalah yang memiliki $IKB = 0,30-0,70$.

Berdasarkan contoh pada Tabel 4, maka $IKB = \frac{20}{40} \times 100\% = 50\% = 0,50$

Berdasarkan kriteria IKB tersebut, maka butir yang memiliki $IKB = 0,50$ termasuk butir *sedang*, tetapi masih ditoleransi untuk dapat digunakan sebagai butir standar.

Apabila jumlah responden relatif banyak, maka Guru menetapkan $KA = KB = 27\%$ dari jumlah responden seluruhnya. Misalkan $N = 100$, maka $KA = KB = 27$, dan menurut formula (1), maka $T = 54$. Sedangkan R tentunya menyesuaikan dengan jumlah siswa dari semua kelompok yang mampu menjawab benar. Dengan demikian, IKB akan dapat dihitung dengan menggunakan formula (1).

7. Hitung Indek Dayabeda Butir (IDB) dengan formula berikut.

$$IDB = \frac{R_{KA} - R_{KB}}{\frac{1}{2}T} \quad (2)$$

dengan $IDB =$ Indeks Dayabeda Butir, $R_{KA} =$ jumlah responden Kelompok Atas yang menjawab benar, $R_{KB} =$ jumlah responden Kelompok Bawah yang menjawab benar, dan $T =$ jumlah responden seluruhnya.

Nilai IDB bergerak dari $-1,00$ s.d $+1,00$. Apabila IDB bernilai positif, butir tersebut memiliki dayabeda yang positif, yang berarti bahwa porsi siswa yang lebih tahu tentang jawaban benar lebih besar dibandingkan dengan porsi siswa yang tidak tahu. Apabila IDB bernilai nol, butir tersebut memiliki dayabeda nol, artinya butir tersebut tidak mampu membedakan antara siswa yang tahu jawaban benar dengan siswa yang tidak tahu. Hal ini terjadi, karena beberapa hal, yaitu: (1) butir terlalu mudah atau terlalu sukar, sehingga mungkin semua siswa salah atau semua siswa benar, (2) butir tersebut membingungkan sebagai akibat konstruksinya ambigu (mungkin menimbulkan penafsiran ganda). Apabila porsi siswa yang tidak tahu jawaban benar lebih banyak dibandingkan dengan yang tahu, maka IDB menjadi negatif. Hal ini bisa terjadi mungkin disebabkan karena konstruksi tes bersifat ambigu, atau kunci jawabannya yang salah. Secara umum, semakin tinggi IDB suatu butir semakin besar kemungkinan butir tersebut mampu membedakan antara siswa yang tahu jawaban benar dengan siswa yang tidak tahu. Kriteria IDB dapat diacu, rentangan berikut, IDB: $0,00-0,20$ adalah *sangat rendah*, $0,20-0,40$ adalah *rendah*, $0,40-0,60$ adalah *sedang*, $0,60-0,80$ adalah *tinggi*, $0,80-1,00$ adalah *sangat tinggi*. Untuk tes standar dianjurkan menggunakan tes yang memiliki $IDB > 0,20$.

Untuk contoh pada Tabel 4, berarti $R_{KA} = 12$, $R_{KB} = 8$, sehingga dapat dihitung $IDB = 0,20$, yang berkategori *rendah*.

8. Menentukan keefektifan pengecoh (*distracters effectiveness*). Kriterianya, adalah pengecoh akan efektif apabila jumlah siswa KB lebih banyak memilih dibandingkan

jumlah siswa KA. Pada contoh Tabel 4, dapat diestimasi bahwa pilihan A dan D adalah *efektif*, sedangkan pilihan B dan E *tidak efektif*.

Analisis Butir untuk Tes Essai

Untuk tes essay, analisis butir hanya menyangkut IKB dan IDB. Prosedur analisisnya adalah sebagai berikut.

1. Lakukan koreksi terhadap semua jawaban responden pada semua butir tes, kemudian tabulasi ke dalam tabel kerja, seperti Tabel 5. Pada Tabel 5, dicontohkan skor maksimum tiap item adalah 5 dan skor minimumnya adalah 0.

Tabel 5 Skor-skor butir tes essai

No. Resp	Skor Perbutir										Skor Total
	1	2	3	4	5	6	7	8	9	10	
1	3	3	4	4	4	5	5	2	1	0	31
2	5	5	5	5	5	5	5	4	4	4	47
3	3	3	3	2	2	2	2	5	4	1	27
4	1	1	1	0	0	0	2	3	4	5	17
5	3	3	3	3	4	5	3	4	2	0	30
6	4	4	4	2	2	3	3	3	1	2	28
7	3	3	2	1	2	2	2	4	5	3	27
8	5	5	5	5	5	4	4	4	4	3	44
9	4	4	4	4	4	4	2	2	2	5	35
.
n

2. Urutkan skor-skor responden tersebut dari yang tertinggi ke yang terendah, seperti pada Tabel 6.

Tabel 6 Skor butir tes essai setelah skor-skor responden diurutkan

No. Resp	Skor Perbutir										Skor Total
	1	2	3	4	5	6	7	8	9	10	
2	5	5	5	5	5	5	5	4	4	4	47
8	5	5	5	5	5	4	4	4	4	3	44
9	4	4	4	4	4	4	2	2	2	5	35
1	3	3	4	4	4	5	5	2	1	0	31
5	3	3	3	3	4	5	3	4	2	0	30
6	4	4	4	2	2	3	3	3	1	2	28
3	3	3	2	1	2	2	2	4	5	3	27
7	3	3	3	2	2	2	2	5	4	1	27
9	1	1	1	0	0	0	2	3	4	5	17
.
n

- Berdasarkan Tabel 6, tetapkan 25% dari urutan nomor 1 ke bawah sebagai KA dan 25% dari urutan terakhir ke atas sebagai KB.
- Hitung jumlah skor-skor untuk masing-masing butir baik pada KA maupun pada KB.
- Tentukan IKB dan IDB masing-masing dengan formula-formula berikut.

$$IKB = \frac{\sum H + \sum L - (2N \times Score_{\min})}{2N(Score_{\max} - Score_{\min})} \quad (3)$$

$$IDB = \frac{\sum H - \sum L}{N(Score_{\max} - Score_{\min})} \quad (4)$$

dengan $\sum H$ = jumlah skor Kelompok Atas (KA), $\sum L$ = jumlah skor Kelompok Bawah (KB), N = jumlah responden pada KA atau KB, $Score_{\max}$ = skor tertinggi butir, dan $Score_{\min}$ = skor terendah butir. Kriteria IKB dan IDB dapat diacu ketentuan yang berlaku pada analisis butir tes pilihan ganda.

Sebagai contoh, misalkan untuk sebuah butir tes yang dijawab oleh 30 responden pada KA dan KB, diperoleh sebaran butir seperti pada Tabel 7.

Tabel 7 Sebaran skor sebuah butir tes dari 30 KA dan KB

Skor Butir (S)	KA		KB	
	f	f × S	f	f × S
5	8	40	3	15
4	5	20	7	28
3	6	18	8	24
2	4	8	6	12
1	7	7	2	2
0	0	0	4	0
Total	30	93	30	81

Berdasarkan data pada tabel 7, dan dengan menggunakan formula (3) dan (4) dapat dihitung IKB dan IDB sebagai berikut.

$$IKB = \frac{93 + 81 - (60 \times 0)}{60(5 - 0)} = \frac{174}{300} = 0,58$$

$$IDB = \frac{93 - 81}{30(5 - 0)} = \frac{12}{150} = 0,08$$

Hasil analisis menunjukkan bahwa butir tes tersebut memiliki IKB dengan kategori *sedang* dan IDB yang berkategori *sangat rendah*.

Menggunakan Hasil-Hasil Analisis Butir

Beberapa nilai yang cukup memberikan mamfaat bagi Guru tentang data analisis butir adalah sebagai berikut: (1) Membantu Guru dalam melakukan justifikasi terhadap manfaat atau kualitas tes, (2) Melakukan revisi terhadap tes yang telah dikembangkan, (3) Mengembangkan tes untuk kebutuhan berikutnya, (4) Meningkatkan keterampilan Guru dalam mengkonstruksi tes, (5) Menyediakan nilai diagnostik dan membantu dalam perencanaan aktivitas belajar berikutnya, (6) Sebagai bahan diskusi tentang hasil-hasil tes, (7) Dapat memberikan pengalaman belajar bagi siswa yang terlibat sebagai asisten.

Walaupun IKB, IDB, dan Keefektifan Pengecoh telah memberikan manfaat yang sangat berarti bagi Guru dalam menjustifikasi kualitas tes yang telah digunakan, namun manfaat yang lebih penting adalah melakukan revisi terhadap tes atau butir-butirnya. Revisi butir tes memiliki nilai lebih ekonomis dibandingkan mengembangkan butir yang baru. Berdasarkan data analisis butir, Guru dapat mengacu pada bagian butir revisi dan menentukan keefektifan perubahan. Revisi ini dilakukan secara berkelanjutan dan proses *check-recheck* dapat mengarahkan peningkatan keterampilan dalam konstruksi tes, utamanya dalam penulisan butir dan menentukan kualitas pengecoh secara teoretik. Satu hal yang sangat penting, bahwa sebelum melakukan perubahan pada butir, seyogyanya terlebih dulu dilakukan koreksi secara cermat pada butir tersebut.

Kegunaan yang lain data analisis butir terkait dengan dampaknya terhadap prosedur pembelajaran. Apabila untuk beberapa butir, semua siswa salah, barangkali pembelajaran terkait dengan materi yang direpresentasi oleh butir tersebut kurang mampu memfasilitasi siswa untuk meningkatkan pemahaman. Kemungkinan lain, bahwa ada materi pelajaran yang telah ditetapkan butir tesnya belum sempat dipelajari oleh siswa. Atas dasar data analisis butir tersebut, Guru akan mampu melakukan refleksi dan sekaligus merinci kembali ketepatan pembelajaran selanjutnya yang secara komprehensif mencakup semua dimensi konseptual yang telah ditetapkan menjadi materi tes.

Data analisis butir juga dapat memberikan manfaat balikan bagi siswa. Beberapa peran balikan yang mampu ditampilakn oleh analisis butir tes adalah, (1) Untuk mengoreksi kesalahan siswa, (2) Memotivasi siswa untuk lebih baik ketika tes berikutnya, (3) Menunjukkan pada siswa bahwa tujuan belajar mereka seyognya diuji ketercapaiannya melalui tes, sehingga siswa mampu melakukan organisasi kebiasaan belajarnya, (4) Memberi peluang kepada siswa untuk mempertanyakan tentang jawaban benar butir tersebut, sehingga mampu meluruskan miskonsepsi atau salah pengertian yang mereka miliki.

Interpretasi data Analisis Butir

Beberapa interpretasi yang dapat ditampilkan terkait dengan data analisis butir, adalah sebagai berikut.

1. *Data analisis butir tidak analog dengan validitas butir.* Tes-tes psikologi seyogyanya memperhitungkan validitas butir, seperti *construct validity*. Namun, untuk tes hasil belajar, meneliti konsistensi internal butir tampak lebih penting dibandingkan menganalisis validitasnya. Hal ini karena tes hasil belajar lebih menyandarkan diri pada validitas isi. Jadi kriteria internal menjadi lebih penting untuk diperhitungkan. Kriteria internal mendasarkan diri pada skor total tes.
2. *Indesk Dayabeda Butir tidak selalu suatu ukuran kualitas butir.* Artinya, rendahnya IDB bukan ukuran rendahnya kualitas butir tersebut. Ada beberapa alasan mengapa IDB suatu butir bisa bernilai rendah. (1) Semakin sukar atau semakin mudah suatu butir, semakin rendah IDB-nya, tetapi Guru sering membutuhkan item-item yang sukar atau mudah agar representatif terhadap karakteristik materi dan tujuan belajar siswa. Jadi butir yang kualitasnya relatif rendah tetap dipertimbangkan sebagai butir tes ketiak butir tersebut memang diperlukan. (2) Tujuan item hubungannya dengan tes keseluruhan akan mempengaruhi besarnya IDB. Hal ini karena skor total merupakan kriteria internal yang digunakan. Skor total merupakan gabungan skor keseluruhan butir, baik yang sukar maupun yang mudah, dari berbagai pokok bahasan dengan segala keragaman karakteristiknya, dan dari keragaman jenjang tes. Misalkan butir pada jenjang *evaluasi* (C6) hanya 5% dari keseluruhan tes, kemungkinan besar butir-butir yang merepresentasikan jenjang kemampuan tersebut akan memiliki IDB yang relatif rendah, tetapi tetap dibutuhkan sebagai butir tes.
3. *Data analisis butir bersifat tentatif.* Sering dipersepsi oleh para Guru bahwa IKB dan IDB bersifat tetap. Hal ini tidak benar. Data analisis butir dipengaruhi oleh karakteristik responden, jumlah responden, pembelajaran yang diterapkan oleh Guru, dan peluang kesalahan. Pertimbangan Guru terhadap kualitas butir hendaknya lebih ditekankan pada mampu atau tidaknya suatu butir mencerminkan pengukuran terhadap tujuan belajar dibandingkan dengan besarnya IKB dan IDB. Jadi, ketika analisis IKB dan IDB yang tidak bisa dilakukan, maka proses pengembangan butir yang lebih baik yang mampu mencerminkan validitas isi yang lebih komprehensif akan menjadi lebih penting dibandingkan dengan hanya memperhitungkan besarnya IKB dan IDB.
4. *Hindari seleksi butir-butir tes secara murni pada basis sifat-sifat statistik butir tersebut.* Satu dari cara-cara terbaik untuk menseleksi butir-butir tes adalah dengan memilih butir-butir yang memiliki IKB dan IDB yang tepat. Ini bukan berarti bahwa analisis statistik menjadi faktor utama, khususnya tes hasil belajar buatan Guru untuk sekolah dasar, di mana sampel mungkin hanya 30 orang, yang boleh jadi tidak reliabel. Alasan lain, bahwa IKB boleh jadi dipengaruhi oleh cara siswa menjawab dengan tebak-tebakan, cara menempatkan jawaban benar, dan cara menempatkan urutan item pada tes tersebut. Alasan lain, bahwa seleksi secara statistik sering memberikan pertimbangan yang bias, bahwa item boleh jadi tidak mengukur tujuan

pembelajaran, yang berarti membuat validitas isi menjadi rendah. Menseleksi butir-butir tes akan lebih baik berdasarkan kategori taksonomi yang secara jelas telah tergambar dalam kisi-kisi tes.

Konsistensi Internal Butir dan Tes

Analisis butir tes standar tidak dapat ditentukan hanya oleh IKB, IDB, dan untuk tes pilihan ganda oleh keefektifan pengecoh, tetapi juga harus ditambah oleh analisis konsistensi internal baik konsistensi internal butir maupun konsistensi internal tes (reliabilitas tes).

Konsistensi internal butir adalah tingkatan konsistensi butir dalam pengukuran apa yang seharusnya diukur. Sedangkan, Gay (1987) menyatakan konsistensi internal butir adalah derajat konsistensi pengukuran yang ditampilkan oleh butir terhadap apa yang ingin diukur. Jadi konsistensi butir berkenaan dengan tingkatan atau derajat yang menunjukkan seberapa jauh butir dapat mengukur secara konsisten apa yang seharusnya diukur.

Konsistensi internal butir dapat diestimasi dari indeks korelasi antara skor butir dan skor total. (Long *et al*, 1985). Indeks korelasi butir-total dapat dihitung dengan formula *product moment*

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}} \quad (5)$$

dengan r_{xy} = indeks korelasi butir-total, N = jumlah responden, X = skor butir, dan Y = skor total. Kriteria estimasi yang digunakan adalah indeks *korelasi butir-total* di atas 0,30 disebut sebagai butir yang memiliki derajat konsistensi internal butir yang tinggi, sedangkan indeks korelasi yang berada pada rentangan 0,10-0,30 direkomendasikan untuk direvisi (Long *et al*, 1985).

Wiersma (1991) menyatakan konsistensi internal tes (reliabilitas tes) berarti “konsistensi dari tes dalam mengukur apa yang seharusnya diukur”. Pengukuran konsisten berarti akan memberikan hasil yang sama untuk subjek yang sama pada waktu yang berbeda. Koefisien reliabilitas tes dapat bernilai antara 0,00-1,00. Gay (1987) menyatakan “reliabilitas tes adalah derajat pada mana suatu tes dapat mengukur secara konsistens apa yang seharusnya diukur”.

Konsistensi internal tes (Reliabilitas internal tes) dapat ditentukan dengan beberapa metode, (1) metode belah dua, (2) metode Kuder-Rechadson 20 (K-R 20), (3) metode K-R 21, dan (4) koefisien alfa Cronbach.

Metode belah dua dapat ditempuh dengan prosedur ganjil-genap. Indeks korelasi tes ganjil-genap dapat dihitung dengan formula *product moment* seperti formula (5), sedangkan indeks korelasi keseluruhan tes dapat dihitung dengan formula Spearman-Brown (Mehrens & lehmann, 1984),

$$r_{xx} = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}} \quad (6)$$

dengan r_{xx} = indeks korelasi keseluruhan tes, $r_{\frac{1}{2}\frac{1}{2}}$ = indeks korelasi setengah dari jumlah tes keseluruhan.

Untuk skor-skor butir yang bersifat dikotomis, maka koefisien reliabilitas dihitung dengan Metode K-R 20 dengan formula (Mehrens & Lehmann, 1984)

$$r_{xy} = \frac{n}{n-1} \left[1 - \frac{\sum pq}{S_x^2} \right] \quad (7)$$

dengan n = jumlah butir tes, p = persentase responden yang menjawab benar, q = persentase responden yang menjawab salah, S_x^2 = varians keseluruhan tes.

Apabila Indeks Kesukaran Butir (IKB) bersifat homogen, yang berarti bahwa p relatif konstan untuk keseluruhan butir, maka indeks reliabilitas tes dihitung dengan metode K-R 21 dengan formula (Mehrens & Lehmann, 1984)

$$r_{xx} = \frac{n}{n-1} \left[1 - \frac{\bar{X}(n - \bar{X})}{nS_x^2} \right] \quad (8)$$

dengan \bar{X} = nilai rata-rata responden. Metode K-R 21 juga digunakan untuk skor butir yang bersifat dikotomis.

Apabila skor butir bersifat non-dikotomis, maka koefisien reliabilitas tes diestimasi berdasarkan koefisien alfa Cronbach. Koefisien alfa Cronbach dapat dihitung dengan formula Mehrens & Lehmann (1984)

$$\text{Alfa Cronbach} = \frac{n}{n-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right] \quad (9)$$

dengan n = jumlah butir tes, S_i^2 = varian butir, dan S_x^2 = varian total tes.

Kriteria yang dapat diacu adalah koefisien reliabilitas $\geq 0,80$ menyatakan tes tersebut *acceptable* (Long *et al*, 1985). Oleh karena koefisien reliabilitas secara wajar bergerak pada interval 0,00-1,00, maka kriteria-kriteria: 0,00-0,20 adalah *sangat rendah*, 0,20-0,40 *rendah*, 0,40-0,60 *sedang*, 0,60-0,80 *tinggi*, dan 0,80-1,00 *sangat tinggi* dapat pula diacu sebagai kriteria penolakan atau penerimaan reliabilitas internal. Tes hasil belajar dengan indek reliabilitas berada pada kategori *sedang*, *tinggi*, dan *sangat tinggi* ditoleransi untuk diterima sebagai perangkat tes yang relatif baku.

Daftar Rujukan

- Gay, L. R. 1987. *Education research, Competencies for analysis and application*. Third edition. Columbus: Merrill Publishing Company.
- Long, T. J., Convey, J. J., & Chwalek, A. R. 1985. *Completing dissertation in the behavioral sciences and education*. London: Jossey-Bass Publishers.
- Mehrens, W. A. & Lehmann, I. J. 1984. *Measurement and evaluation in education and psychology*, Third edition. New York: Holt, Rinehart and Winston.
- Wiersma, W. 1991. *Research methods in education*. Fifth edition. Boston: Allyn and Bacon.