



BAB IV

VALIDITAS DAN RELIABILITAS

Prof.Dr.Patta Bundu,M.Ed.

A. VALIDITAS TES

Pada dasarnya, istilah validitas berasal dari kata *validity* (kesahihan)¹ yang merujuk pada ketepatan instrumen mengukur aspek-aspek materi ajar atau aspek-aspek perilaku yang seharusnya diukur. Instrumen yang valid berarti alat ukur yang digunakan untuk mengumpulkan data atau informasi itu valid. Dalam kegiatan sehari-hari, kita dapat mengatakan meteran adalah alat ukur yang valid untuk mengukur panjang/ jarak, tetapi tidak valid untuk mengukur berat sebuah benda. Demikian pula, stopwatch atau arloji dikatakan valid sebagai alat ukur waktu tetapi tidak valid untuk mengukur suhu. Dalam proses pembelajaran istilah validitas dapat diartikan sebagai ketepatan suatu tes dalam menghasilkan data atau informasi yang sesuai dengan tujuan atau keputusan yang akan dibuat.

Untuk menentukan apakah suatu tes hasil belajar telah memiliki validitas atau ketepatan mengukur dapat dianalisis dari

¹ Dalam buku ini yang digunakan adalah *valid* dan *validitas*

segi tes itu sendiri secara totalitas atau dari segi butir-butir soal yang menyusun tes tersebut. Ditinjau dari segi penganalisisan tes hasil belajar, dapat dilakukan dengan dua cara. *Pertama*, penganalisaan dilakukan dengan menggunakan logika (*logical validity*) atau validitas rasional. Validitas rasional biasa disebut juga dengan validitas internal. Kedua penganalisisan yang dilakukan berdasarkan kenyataan empiris (*emperical validity*) atau validitas eksternal.

1. Validitas Rasional

Validitas rasional adalah validitas yang diperoleh atas dasar pemikiran secara logis. Validitas rasional biasa juga disebut validitas internal. Untuk menentukan apakah instrumen yang digunakan sudah valid dari segi rasio dapat ditinjau dari dua segi yakni isinya (*content*) dan segi konstruksinya (*construct*).

- 1) Validitas Isi (*Content validity*). Validitas isi adalah validitas yang diperoleh setelah dilakukan penganalisisan, penelusuran dan pengujian terhadap isi yang terkandung dalam instrumen hasil belajar. Validitas ini adalah validitas yang ditilik dari segi isi instrumen yakni isinya telah dapat mewakili keseluruhan materi atau bahan pelajaran yang seharusnya diujikan.
- 2) Validitas Konstruksi (*Construct validity*). Validitas konstruksi atau validitas bangun adalah validitas ditilik dari susunan, kerangka, dan aspek yang diukur oleh suatu instrumen asesmen. Validitas bangun dapat dianalisis dengan mencocokkan aspek yang ingin diukur dengan susunan setiap butir instrumen.

2. Validitas Empirik

Validitas empirik adalah ketepatan mengukur yang didasarkan pada hasil analisis yang bersifat empirik yang bersumber dari hasil pengamatan di lapangan, sehingga validitas empirik disebut juga validitas eksternal atau validitas lapangan. Disebut validitas eksternal karena kriteria validitasnya dihubungkan dengan instrumen lain yang dianggap sebagai kriterium/ patokan. Untuk menentukan apakah instrumen asesmen hasil belajar memiliki validitas empirik dapat dilihat dari dua aspek yakni ketepatan instrumen memprediksi / meramal hasil yang akan dicapai (*predictive validity*) dan keakuratannya sebagai pembanding pada kriteria tertentu (*concurrent validity*).

- a) Validitas Ramalan (*predictive validity*). Yang dimaksud dengan validitas prediksi adalah kondisi yang menunjukkan seberapa jauh suatu instrumen mampu meramalkan apa yang bakal terjadi di masa datang.
- b) Validitas Bandingan (*concurrent validity*). Suatu instrumen memiliki validitas bandingan jika instrumen tersebut dalam kurun waktu yang sama dengan tepat mampu menunjukkan adanya hubungan yang searah antara instrumen pertama dengan yang berikutnya. Untuk mengetahui ada tidaknya hubungan antara instrumen pertama dengan yang berikutnya dapat dianalisis secara statistik dengan teknik analisis *product moment* dari Karl Pearson.

$$\text{Pearson } r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}} \cdot \sqrt{\{N\sum Y^2 - (\sum Y)^2\}}}$$

- r = koefisien korelasi Pearson
- $\sum XY$ = jumlah hasil kali skor X dan Y
- $\sum X$ = jumlah skor X
- $\sum Y$ = jumlah skor Y
- $\sum X^2$ = jumlah kuadrat skor X
- $\sum Y^2$ = jumlah kuadrat skor Y
- N = jumlah peserta (pasangan skor)

Jika korelasi antara variabel X (tes pertama) dengan variabel Y (tes berikutnya), positif dan cukup berarti (signifikan) maka tes tersebut memiliki validitas bandingan. Misalnya, kita ingin menghitung validitas bandingan antara dua hasil tes terhadap 10 orang peserta didik dengan data sebagai berikut.

Tes I (X) : 28, 24, 23, 23, 22, 20, 19, 19, 17, 15

Tes II (Y) : 50, 58, 64, 47, 70, 44, 53, 48, 41, 52

Untuk memudahkan perhitungannya kita susun data dalam tabel, lalu menghitung semua data yang diperlukan dalam menggunakan koefisien korelasi pearson.

Tabel 3.3. Persiapan Korelasi *Pearson Product Moment*

N	X	X ²	Y	Y ²	XY
1	28	784	50	2500	1400
2	24	576	58	3364	1392
3	23	529	64	4096	1472
4	23	529	47	2209	1081
5	22	484	70	4900	1540
6	20	400	44	1936	880
7	19	361	53	2809	1007
8	19	361	48	2304	912
9	17	289	41	1681	697
10	15	225	52	2704	780
	ΣX= 210 M = 21	ΣX ² = 4538	ΣY= 527 N = 10 M = 52,7	ΣY ² = 28503	ΣXY = 11161

Menghitung koefisien korelasi dengan rumus:

$$Pearson r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}} \cdot \sqrt{\{N\Sigma Y^2 - (\Sigma Y)^2\}}}$$

$$Pearson r = \frac{10.11161 - (210)(257)}{\sqrt{\{10.4538 - (210)^2\}} \cdot \sqrt{\{10.28503 - (257)^2\}}}$$

$$Pearson r = \frac{111610 - 110670}{\sqrt{\{45380 - 44100\}} \cdot \sqrt{\{285030 - 277729\}}}$$

$$Pearson r = \frac{940}{\sqrt{\{1280\}} \cdot \sqrt{\{7301\}}}$$

$$Pearson r = \frac{940}{35,78 \cdot 85,44}$$

$$Pearson r = 0,307$$

. Pada tabel untuk $N = 10$ dengan taraf kepercayaan (signifikansi) $0,01 = 0,7079$, dan pada taraf $0,05 = 0,5760$. Kita mengatakan bahwa validitas bandingannya tidak signifikan karena koefisien korelasi yang diperoleh ($0,3075$) < dari nilai tabel. Pada dasarnya koefisien korelasi bergerak dari -1 sampai $+1$. Makin tinggi koefisien korelasi makin signifikan hubungan kedua variabel yang dibandingkan.

a. Validitas Butir Instrumen

Validitas butir instrumen adalah ketepatan mengukur yang dimiliki oleh satu butir soal dalam mengukur apa yang seharusnya diukur oleh butir soal tersebut. Hubungan antara butir-butir soal dengan instrumen asesmen hasil belajar secara keseluruhan sangat erat, karena instrumen yang baik akan ditentukan tingkat validitas butir-butir soal yang menyusun instrumen tersebut. Hal ini memberikan pengertian pula bahwa validitas dari masing-masing butir yang membangun instrumen asesmen akan dapat diketahui dengan melihat besar kecilnya dukungan masing-masing butir terhadap validitas instrumen asesmen secara keseluruhan.

Eratnya hubungan antara validitas butir dengan validitas instrumen secara keseluruhan maka satu butir instrumen dikatakan memiliki validitas tinggi jika skor-skor pada butir tersebut memiliki kesesuaian arah atau berkorelasi positif yang berarti (*significance positive*) dengan skor total instrumen. Dengan kata lain, sebuah soal dikatakan valid jika skor butir tersebut mempunyai korelasi positif yang signifikan dengan skor totalnya. Oleh karena itu para pakar asesmen menyarankan penggunaan analisis korelasi secara statistik untuk mengukur validitas butir instrumen. Pertanyaan yang mungkin timbul ialah teknik korelasi apa yang cocok digunakan untuk menguji validitas butir instrumen? Pemilihan teknik korelasi ini bergantung pula pada jenis skor yang akan dianalisis. Pada butir-butir instrumen dengan bentuk soal obyektif pemberian skor dapat dilakukan dengan memberi skor 1 (satu) untuk jawaban yang benar dan 0 (nol) untuk jawaban yang salah. Dalam istilah statistik data sejenis ini disebut data diskrit murni atau data dikotomik, sedangkan skor totalnya berupa data kontinu (seperti skor 3, 7, 5, 9, dan seterusnya). Oleh para pakar statistik menyarankan penggunaan teknik korelasi "*point biserial*" untuk data seperti itu. Apabila butir-butir instrumen berupa tes uraian, atau skala maka skornya tentu bukan 1 (satu) atau 0 (nol) tetapi merupakan rentang dengan skor mungkin 1, 2, 3, atau 4 dan seterusnya. Untuk data seperti ini disarankan penggunaan teknik korelasi "*Pearson product moment*".

1). Perhitungan Validitas Butir dengan Korelasi "Point Biserial"

Teknik korelasi point biserial mempunyai pola rumus :

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

r_{pbis} = koefisien korelasi point biserial

M_p = skor rata-rata hitung untuk butir yang dijawab

betul

M_t = skor rata-rata dari skor total

Sd_t = standar deviasi skor total

p = proporsi peserta didik yang menjawab betul pada butir yang diuji

q = proporsi peserta didik yang menjawab salah pada butir yang diuji

Misalkan setelah memeriksa pekerjaan peserta didik pada butir-butir soal obyektif dan memberikan skor (skor 1 untuk jawaban benar dan skor 0 untuk jawaban salah), kita memiliki data dan ditabulasi sebagai berikut

Tabel 3.4. Tabulasi Skor Butir Soal Bentuk Obyektif

No Uru t	Nomor Butir Instrumen											
	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9	1 0	1 1	1 2
1	0	1	1	1	0	0	0	0	0	1	0	0
2	0	0	1	1	1	1	1	0	1	1	0	0
3	1	0	0	0	0	0	1	1	0	0	1	1
4	0	0	0	1	0	0	0	0	1	0	1	0
5	1	1	1	1	0	1	0	0	0	1	1	1

6	0	0	0	1	1	0	0	0	1	1	0	0
7	0	0	1	1	0	1	1	1	0	0	0	1
8	0	1	1	0	0	1	1	0	1	0	0	1
9	0	1	1	0	0	0	1	0	1	0	0	1
10	0	1	1	0	0	0	1	0	1	0	0	1
11	1	1	1	1	0	1	0	0	1	1	1	1
12	1	0	1	1	1	1	1	0	1	1	1	1
13	1	1	0	0	1	1	1	1	1	0	1	1
14	1	0	0	1	1	1	1	1	1	0	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1

Untuk menghitung validitas setiap butir instrumen pada data diatas dapat dilakukan beberapa langkah:

- 1) Menghitung jumlah peserta didik yang menjawab benar pada setiap butir dan jumlah jawaban yang benar untuk setiap peserta didik.
- 2) Dari data pada langkah pertama akan dapat dihitung nilai p dan q, jumlah skor total ($\sum X_t$) dan jumlah kuadrat skor total ($\sum X_t^2$).
- 3) Menghitung rata-rata skor total
- 4) Menghitung standar deviasi total
- 5) Menghitung M_p (rata-rata hitung skor total yang dijawab betul).
- 6) Menghitung validitas butir dengan point beserial

Penyelesaian:

- a. Membuat Tabel Perhitungan untuk skor total dan kuadrat skor total.

Tabel. 3.5. Persiapan Perhitungan Validitas (Korelasi Point Biserial)

No	Nomor Butir Instrumen												X _t	X _t ²
	1	2	3	4	5	6	7	8	9	10	11	12		
1	0	1	1	1	0	0	0	0	0	1	0	0	4	16
2	0	0	1	1	1	1	1	0	1	1	0	0	7	49
3	1	0	0	0	0	0	1	1	0	0	1	1	5	25
4	0	0	0	1	0	0	0	0	1	0	1	0	3	9
5	1	1	1	1	0	1	0	0	0	1	1	1	8	64
6	0	0	0	1	1	0	0	0	1	1	0	0	4	16
7	0	0	1	1	0	1	1	1	0	0	0	1	6	36
8	0	1	1	0	0	1	1	0	1	0	0	1	6	36
9	0	1	1	0	0	0	1	0	1	0	0	1	5	25
10	0	1	1	0	0	0	1	0	1	0	0	1	5	25
11	1	1	1	1	0	1	0	0	1	1	0	1	8	64
12	1	0	1	1	1	1	1	0	1	1	0	1	9	81
13	1	1	0	0	1	1	1	1	1	0	0	1	8	64
14	1	0	0	0	1	1	1	1	1	0	0	1	7	49
15	1	1	1	1	1	1	1	1	1	1	1	1	12	144
N	7	8	10	9	6	9	10	5	11	7	4	11	97	703
P	0,47													
Q	0,53													

$$\Sigma X_t = 97$$

$$\Sigma X_t^2 = 703$$

Nilai p = jumlah yang menjawab benar pada butir tertentu dibagi jumlah peserta didik (pada butir 1, misalnya, yang menjawab benar 7 orang, berarti $p = 7/15 = 0,47$)

$$q = 1 - p \text{ (pada butir 1; } q = 1 - 0,47 = 0,53)$$

Demikian seterusnya sehingga didapatkan nilai p dan q.

b. Menghitung rata-rata skor total

$$M_t = \frac{\sum X_t}{N} = \frac{97}{15} = 6,46$$

- c. Menghitung standar deviasi total dengan menggunakan rumus²:

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

$$Sd_t = \sqrt{\frac{703}{15} - \left(\frac{97}{15}\right)^2}$$

$$Sd_t = \sqrt{46,87 - 41,82} = 2,247$$

- d. Menghitung M_p setiap butir soal (rata-rata hitung dari skor total yang dijawab dengan betul). Perhatikan cara menghitungnya:
Butir 1. Jumlah yang menjawab betul 7 orang (peserta didik No. 3, 5, 11, 12, 13, 14, 15), sedangkan skor total setiap peserta didik: $5 + 8 + 8 + 9 + 8 + 7 + 12 = 57$.

$$M_p = 57/7 = 8,14$$

Perhatikan contoh berikut, misalnya untuk butir 10. Jumlah yang menjawab benar juga 7 orang (peserta didik No. 1, 2, 5, 6, 11, 12, dan 15). Skor total setiap peserta didik adalah: $4 + 7 + 8 + 4 + 8 + 9 + 12 = 52$

$$M_p = 52/7 = 7,43$$

² Banyak rumus Standar Deviasi yang digunakan. Yang digunakan pada buku ini adalah menghitung standar deviasi langsung dari skor mentah (raw scores).

Ternyata meskipun yang menjawab benar pada butir tertentu sama jumlahnya, tetapi M_p tidak sama, karena besarnya M_p ditentukan pula oleh skor total peserta didik yang menjawab benar pada butir tersebut.

Secara lengkap dapat disajikan seperti pada tabel di bawah

Tabel 3.6. Hasil Perhitungan M_p Setiap Butir Instrumen

No. Butir	Peserta didik yang menjawab betul	Skor total yang menjawab betul	M_p
1	3,5,11,12,13,14,15	$5+8+8+9+8+7+12 = 57$	$57/7 = 8,14$
2	1,5,8,9,10,11,13,15	$4+8+6+5+5+8+8+12 = 56$	$56/8 = 7,00$
3	1,2,5,7,8,9,10,11,12,15	$4+7+8+6+6+5+5+8+9+12 = 70$	$70/10 = 7,00$
4	1,2,4,5,6,7,11,12,15	$4+7+3+8+4+6+8+9+12 = 61$	$61/9 = 6,78$
5	2,6,12,13,14,15	$7+4+9+8+7+12 = 47$	$47/6 = 7,83$
6	2,5,7,8,11,12,13,14,15	$7+8+6+6+8+9+8+7+12 = 71$	$71/9 = 7,89$
7	2,3,7,8,9,10,12,13,14,15	$7+5+6+6+5+5+9+8+7+12 = 70$	$70/10 = 7,00$
8	3,7,13,14,15	$5+6+8+7+12 = 38$	$38/5 = 7,60$
9	2,4,6,8,9,10,11,12,13,14,15	$7+3+4+6+5+5+8+9+8+7+12=74$	$74/11 = 6,73$
10	1,2,5,6,11,12,15,	$4+7+8+4+8+9+12 = 52$	$52/7 = 7,43$
11	3,4,5,15	$5+3+8+12 = 28$	$28/4 = 7,00$
12	3,5,7,8,9,10,11,12,13,14,15	$5+8+6+6+5+5+8+9+8+7+12 = 79$	$79/11 = 7,18$

e. Menghitung validitas butir dengan korelasi point biserial.

Validitas butir 1

$$r_{pbis} = \frac{M_p - M_t}{Sd_t} \sqrt{\frac{p}{q}}$$
$$r_{pbis} = \frac{8,14 - 6,46}{2,247} \sqrt{\frac{0,47}{0,53}} = \mathbf{0,704}$$

Validitas butir 2.

$$r_{pbis} = \frac{M_p - M_t}{Sd_t} \sqrt{\frac{p}{q}}$$
$$r_{pbis} = \frac{7,00 - 6,46}{2,247} \sqrt{\frac{0,53}{0,47}} = \mathbf{0,255}$$

Dengan cara yang sama dapat dihitung r_{pbis} butir yang lainnya.

2) Perhitungan Validitas dengan Korelasi Peasron Product Moment

Seperti telah diuraikan bahwa korelasi point biserial hanya bisa digunakan untuk butir soal bentuk obyektif atau skor-skor dengan data diskrit. Untuk data kontinue atau interval digunakan korelasi *Pearson Product Moment*, dengan formula sebagai berikut:

$$Pearson\ r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}} \cdot \sqrt{\{N\sum Y^2 - (\sum Y)^2\}}}$$

- r = koefisien korelasi Pearson
- $\sum XY$ = jumlah hasil kali skor X dan Y
- $\sum X$ = jumlah skor X
- $\sum Y$ = jumlah skor Y
- $\sum X^2$ = jumlah kuadrat skor X
- $\sum Y^2$ = jumlah kuadrat skor Y
- N = jumlah peserta (pasangan skor)

Misalnya, kita mempunyai data seperti pada tabel di bawah ini.

Tabel 3.7. Data Skor Hasil Belajar dengan Skala 1 – 5.

No	NAMA Peserta didik	Nomor Butir Instrumen									
		01	02	03	04	05	06	07	08	09	10
1	A	5	4	3	3	2	5	5	4	3	4
2	B	5	5	5	3	2	5	4	4	4	4
3	C	4	4	5	4	2	5	4	5	4	4
4	D	4	4	5	4	3	5	3	5	5	5
5	E	5	4	4	5	3	4	3	4	5	5
6	F	3	5	2	5	4	4	4	4	5	5
7	G	3	5	2	5	4	5	4	5	4	5
8	H	2	3	5	5	5	5	5	5	3	4
9	M	4	3	5	4	5	4	5	5	3	5
10	N	2	4	3	2	3	4	3	4	4	4

Penyelesaian:

Langkah pertama: Menghitung jumlah skor total setiap butir instrumen, ($\sum X_i$) dan jumlah skor total semua jawaban peserta didik ($\sum Y$).

Tabel 3.8. Persiapan Perhitungan Korelasi Product Moment

Peserta didik	Nomor Butir Instrumen										Y _t	Y _t ²
	01	02	03	04	05	06	07	08	09	10		
A	5	4	3	3	2	5	5	4	3	4	38	1444
B	5	5	5	3	2	5	4	4	4	4	41	1681
C	4	4	5	4	2	5	4	5	4	4	41	1681
D	4	4	5	4	3	5	3	5	5	5	43	1849
E	5	4	4	5	3	4	3	4	5	5	42	1764
F	3	5	2	5	4	4	4	4	5	5	41	1681
G	3	5	2	5	4	5	4	5	4	5	42	1764
H	2	3	5	5	5	5	5	5	3	4	42	1764
M	4	3	5	4	5	4	5	5	3	5	43	1849
N	2	4	3	2	3	4	3	4	4	4	33	1089
10	37	41	39	40	33	46	40	45	40	45	406	16566

Langkah kedua: Menghitung validitas butir 1

Tabel 3.9. Persiapan Perhitungan Validitas Butir Instrumen

Peserta didik	X ₁	X ₁ ²	Y	Y ²	X ₁ Y
1	5	25	38	1444	190
2	5	25	41	1681	205
3	4	16	41	1681	164
4	4	16	43	1849	172

5	5	25	42	1764	210
6	3	9	41	1681	123
7	3	9	42	1764	126
8	2	4	42	1764	84
9	4	16	43	1849	172
10	2	4	33	1089	66
N = 10	ΣX₁ = 37	ΣX₁² = 149	ΣY = 406	ΣY² = 16566	ΣX₁Y = 1512

Menghitung koefisien korelasi dengan rumus:

$$Pearson r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\}} \cdot \sqrt{\{N \sum Y^2 - (\sum Y)^2\}}}$$

$$Pearson r = \frac{10.1512 - (37)(406)}{\sqrt{\{10.149 - (37)^2\}} \cdot \sqrt{\{10.16566 - (406)^2\}}}$$

$$Pearson r = \frac{15120 - 15022}{\sqrt{\{1490 - 1369\}} \cdot \sqrt{\{165660 - 164836\}}}$$

$$Pearson r = \frac{98}{\sqrt{\{1490 - 1369\}} \cdot \sqrt{\{165660 - 164836\}}}$$

$$Pearson r = \frac{98}{\sqrt{121} \cdot \sqrt{824}}$$

$$Pearson r = \frac{98}{11 \cdot 28,71}$$

$$Pearson r = \mathbf{0,31}$$

Dengan cara yang sama hitunglah validitas butir selanjutnya.

Hasil koefisien korelasi yang diperoleh menunjukkan tingkat validitas butir instrumen yang dianalisis. Makin tinggi koefisien korelasi yang dimiliki makin valid butir instrumen tersebut. Secara umum, jika koefisien korelasi sudah lebih besar dari 0,3 maka butir instrumen tersebut sudah dikategorikan valid (Weiresma and Jurs, 1990). Cara lain, dengan mencocokkan koefisien korelasi yang diperoleh dengan nilai tabel yang tersedia (terlampir) sesuai dengan jumlah peserta didik/ responden yang mengerjakan instrumen yang dianalisis. Jika koefisien korelasi yang diperoleh (r_{hitung}) lebih besar dari nilai tabel (r_{tabel}) maka butir tersebut valid.

f. **Reliabilitas Instrumen**

Reliabilitas biasa juga disebut keterandalan atau kemantapan suatu instrumen yakni sejauh mana instrumen mampu menghasilkan skor-skor hasil asesmen yang stabil dan konsisten. Secara umum ada empat cara yang sering digunakan untuk mengetahui indeks reliabilitas instrumen, yaitu metode tes ulang (*test – retest*), metode belah dua (*split half*), metode paralel (*parallel form*), dan metode Kuder-Richardson²⁰⁻²¹.

1). Metode tes ulang (*test-retest*)

Dalam metode ini digunakan satu jenis asesmen yang dilakukan sebanyak dua kali terhadap kelompok peserta didik yang sama. Dengan kata lain, pelaksanaan tes dilakukan dua kali dalam rentang waktu yang tidak lama. Skor-skor yang didapat pada tes pertama dikorelasikan dengan skor-skor yang didapat

pada tes kedua. Indeks reliabilitas yang diperoleh dengan cara ini disebut koefisien stabilitas (*coefficient of stability*).

2). Metode belah dua / paruhan (*split half*).

Dalam metode ini pelaksanaan tes hanya satu kali tetapi skor-skor butir yang diperoleh dibelah dua, misalnya satu kelompok skor pada butir dengan nomor ganjil dan satu kelompok skor nomor genap. Indeks reliabilitas ditentukan oleh besarnya hubungan antara skor pada butir-butir ganjil dengan skor pada butir-butir genap. Indeks reliabilitasnya disebut koefisien konsistensi internal (*coefficient internal consistency*).

3). Metode bentuk paralel (*parallel form*).

Pengujian dengan metode ini menggunakan dua perangkat tes yang disusun dari ruang lingkup materi yang sama. Bentuk tes mungkin agak berbeda tetapi mengungkap tujuan pembelajaran yang sama. Pelaksanaannya diberikan kepada peserta didik pada waktu yang sama. Skor-skor yang diperoleh dari format tes I dikorelasikan dengan skor-skor yang didapat dari format tes II. Indeks reliabilitas yang diperoleh dengan cara seperti ini disebut koefisien ekuivalensi (*coefficient of equivalence*).

4). Metode *Kuder-Richardson* (KR_{20-21})

Test-retest bisa menimbulkan bias karena tes yang dilakukan merupakan pengulangan untuk tes yang sama. Split half dan paralel kendalanya adalah sukar menentukan pasangan soal yang sama misalnya pada pembagian antara nomor genap dan ganjil atau pada penyusunan pasangan soal yang seimbang. Untuk mengatasi hal tersebut Kuder-Richardson mengembang-

kan teknik lain yaitu mengkorelasikan skor-skor setiap butir soal dengan skor total keseluruhan instrumen.

Metode ini digunakan untuk data *diskrit*. KR_{20} cocok digunakan untuk bentuk tes obyektif. Adapun formula rumus KR_{20} adalah:

$$rKR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum pq}{s^2} \right)$$

rKR_{20} = Koefisien korelasi dengan KR_{20}

k = jumlah butir soal

p = proporsi jawaban benar pada butir tertentu

q = proporsi jawaban salah pada butir tertentu ($q = 1 - p$)

s^2 = varians skor total

Untuk menghitung harga p dan q sama dengan yang telah diuraikan pada pembahasan validitas butir instrumen. Sedangkan s^2 (varians):

$$s^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

Sebagai contoh perhatikanlah data hasil asesmen berikut ini.

Tabel 3.10. Data Hasil Asesmen Bidang Studi IPS-Kelas V SD

No	Nomor Butir Instrumen										X_t	X_t^2
	1	2	3	4	5	6	7	8	9	10		
1	0	1	1	1	0	0	0	1	0	1		
2	0	0	1	1	1	1	1	0	1	1		
3	1	1	1	1	0	1	0	0	1	1		

4	1	0	1	1	1	1	1	0	1	1		
5	1	1	1	1	0	1	0	0	0	1		
6	0	0	0	1	1	0	0	0	1	1		
7	0	0	1	1	0	1	1	1	0	0		
8	0	1	1	0	0	1	1	0	1	0		
9	0	1	1	0	0	0	1	0	1	0		
10	0	1	1	0	0	0	1	0	1	0		

Langkah penyelesaiannya adalah:

- 1) Lengkapi nilai: jumlah (j), p , q , X_t , dan X_t^2

Tabel 3.11. Data Hasil Asesmen Bidang Studi IPS-Kelas V SD

No Urt	Nonor Butir										X_t	X_t^2
	01	02	03	04	05	06	07	08	09	10		
1	0	1	1	1	0	0	0	1	0	1	5	25
2	0	0	1	1	1	1	1	0	1	1	7	49
3	1	1	1	1	0	1	0	0	1	1	7	49
4	1	0	1	1	1	1	1	0	1	1	8	64
5	1	1	1	1	0	1	0	0	0	1	6	36
6	0	0	0	1	1	0	0	0	1	1	4	16
7	0	0	1	1	0	1	1	1	1	0	6	36
8	0	1	1	0	0	1	1	0	1	0	5	25
9	0	1	1	0	0	0	1	0	1	0	4	16
10	0	1	1	0	0	0	1	0	1	0	4	16
J	3	6	9	7	3	6	6	2	8	6	56	332
p	0,3	0,6	0,9	0,7	0,3	0,6	0,6	0,2	0,8	0,6		
q	0,7	0,4	0,1	0,3	0,7	0,4	0,4	0,8	0,2	0,4		
Σpq	0,21	0,24	0,09	0,21	0,21	0,24	0,24	0,16	0,16	0,24	2,00	

- Skor rata-rata total

$$X = \frac{\sum X_t}{N} = \frac{56}{10} = 5,6$$

- Menghitung varians total.

$$s^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

$$s^2 = \frac{332}{10} - 5,6^2 = 1,84$$

4). Menghitung reliabilitas dengan KR_{20} :

$$rKR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum pq}{s^2} \right)$$

$$rKR_{20} = \frac{10}{10-1} \left(1 - \frac{2}{1,84} \right)$$

$$rKR_{20} = 1,11 (- 0,087) = - \mathbf{0,097}$$

Koefisien korelasi berada antara 0 – 1. Suatu instrumen asesmen dikatakan reliabel jika koefisien korelasinya $\geq 0,6$, makin tinggi koefisien korelasi makin reliabel instrumen tersebut, dan sebaliknya. Instrumen asesmen yang kita jadikan contoh termasuk kategori instrumen yang kurang baik karena tidak reliabel. Untuk menjadikan instrumen tersebut reliabel maka validitas butir yang tidak baik dibuang atau diganti. Instrumen yang valid pasti reliabel, tetapi instrumen yang reliabel belum tentu valid.

