

## **BAB III**

### **BENTUK RANCANGAN ANALISIS TES**

#### **3.1. Validitas Tes**

##### **3.1.1. Pengertian Validitas**

Di dalam buku *Encyclopedia of Educational Evaluation* yang ditulis oleh Scarvia B. Anderson dkk disebutkan :

*A test is valid if it measures what it purpose to measure.* Atau jika diartikan lebih kurang demikian: sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur. Dalam bahasa Indonesia “valid” disebut “shahih”.

Menurut Anas Sudjiono (2007:93), suatu tes hasil belajar dapat dikatakan baik jika memiliki ciri atau mempunyai sifat valid atau shahih atau memiliki validitas. Kata valid sering diartikan dengan tepat, benar, shahih, absah; jadi kata validitas dapat diartikan dengan ketepatan, kebenaran, keshahihan, atau keabsahan. Apabila kata valid dikaitkan dengan fungsi tes sebagai alat pengukur, maka sebuah tes dikatakan valid jika tes tersebut dengan secara tepat, secara benar, secara shahih, atau secara absah dapat mengukur apa yang seharusnya diukur.

Dengan demikian alat – alat evaluasi, khususnya tes hasil belajar dapat dikatakan tes yang valid apabila tes tersebut betul – betul dapat mengukur hasil belajar. Jadi bukan sekedar mengukur daya ingatan atau kemampuan bahasa saja.

### 3.1.2. Macam – Macam Validitas

Validitas sebuah tes dapat diketahui dari hasil pemikiran dan dari hasil pengalaman. Dua hal yang pertama akan diperoleh validitas logis (*logical validity*) dan hal yang kedua diperoleh validitas empiris (*empirical validity*). Dua hal inilah yang dijadikan dasar pengelompokan validitas.

Secara garis besar ada dua macam validitas, yaitu validitas tes dan validitas butir (Sudjiono, 2003:65).

#### a. Validitas Tes

Adapun jenis validitas tes secara umum dapat dikelompokkan ke dalam 2 pengelompokkan, yaitu : Validitas Logis dan Validitas Empiris.

##### 1) Validitas Logis

Validitas logis mengandung arti logika / penalaran. Dengan demikian maka validitas logis untuk sebuah instrumen evaluasi menunjuk pada kondisi bagi sebuah instrumen yang memenuhi persyaratan valid berdasarkan hasil penalaran dan sudah dirancang secara baik, sesuai dengan teori dan ketentuan yang berlaku.

Tes hasil belajar yang setelah dilakukan penganalisisan secara rasional ternyata memiliki daya ketepatan mengukur, disebut tes hasil belajar yang telah memiliki validitas logika (*logical validity*).

Ada dua macam validitas logis yang dapat dicapai oleh sebuah instrumen, yaitu : validitas isi dan validitas konstruksi.

a) Validitas Isi

Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan.

Validitas isi dapat diusahakan tercapai sejak saat penyusunan dengan cara memerinci materi kurikulum atau materi buku pelajaran. Suatu tes hasil belajar dikatakan valid, apabila materi tes tersebut betul – betul merupakan bahan-bahan yang representative terhadap bahan – bahan pelajaran yang diberikan.

Dan validitas isi mempersoalkan apakah isi butir tes yang diujikan itu mencerminkan isi materi kurikulum yang seharusnya diukur atau tidak.

b) Validitas Konstruksi

Sebuah tes dikatakan memiliki validitas konstruksi apabila butir – butir soal yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam Tujuan Instruksional Khusus (TIK).

“Konstruksi” dalam pengertian ini bukanlah “susunan” seperti yang sering dijumpai dalam teknik, tetapi merupakan rekaan psikologis, yaitu suatu rekaan yang dibuat para ahli ilmu jiwa yang dengan suatu cara tertentu memerinci isi jiwa atas beberapa aspek seperti ingatan, pemahaman, aplikasi, dan seterusnya.

## 2) Validitas Empiris

Dimaksud dengan validitas empiris adalah memiliki pengertian pengalaman, sehingga sebuah instrument dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Dengan demikian validitas empiris tidak dapat diperoleh hanya dengan jalan menyusun instrument berdasarkan ketentuan seperti halnya validitas logis, tetapi harus dibuktikan dengan hasil analisis yang dilakukan terhadap data hasil pengamatan dilapangan, terbukti bahwa tes hasil belajar itu dengan secara tepat telah dapat mengukur hasil belajar yang seharusnya diukur.

### **b. Validitas Butir**

Menurut Anas Sudjiono (2003:182), yang dimaksud dengan validitas butir dari suatu tes adalah, ketepatan mengukur yang dimiliki oleh sebutir soal (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir soal tersebut. Apabila kita perhatikan secara cermat, maka tes – tes hasil belajar yang dibuat atau disusun oleh guru atau para pengajar sebenarnya adalah merupakan kumpulan dari sekian banyak butir – butir soal; dengan butir mana para penyusun tes ingin mengukur atau mengungkap hasil belajar yang telah dicapai oleh masing – masing individu peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu

tertentu. Eratnya hubungan antara butir soal dengan tes hasil belajar sebagai suatu totalitas itu kiranya dapat dipahami dari kenyataan, bahwa semakin banyak butir – butir item yang dapat di jawab dengan betul oleh testee, maka skor-skor total hasil tes tes tersebut akan semakin tinggi. Sebaliknya semakin sedikit butir-butir item yang dapat dijawab betul oleh testee, maka skorskor total hasil tes itu akan semakin rendah atau semakin menurun.

### 3.1.3. Uji Validitas

Mengukur validitas butir soal dalam penelitian ini, penulis menggunakan salah satu rumus pendekatan uji validitas yang menggunakan **rumus korelasi point biserial**. Arikunto (2006:283) mengemukakan bahwa untuk menganalisis item soal tes maka korelasi point biserial dapat digunakan untuk mencari korelasi untuk mencari korelasi item dengan seluruh tes yang mencari validitas item.

Berikut erupakan rumus point biserial dan tabel korelasi point biserial yang ditunjukkan oleh tabel 3.1 (Arikunto, 2006:283) :

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

#### **Keterangan:**

$r_{pbis}$  = Koefisien Korelasi Point Biserial

$S$  = Standar Deviasi

$M_p$  = Rata – rata jawaban benar

$M_t$  = Rata – rata skor total

$p$  = proporsi jawaban benar terhadap seluruh jawaban

$q$  =  $1 - p$

Koefisien Korelasi ( $r$ )	Tafsiran
$0,40 \leq r < 1,00$	Soal Baik
$0,30 \leq r < 0,40$	Terima dan Perbaiki
$0,20 \leq r < 0,30$	Soal Diperbaiki
$0,00 \leq r < 0,20$	Soal Ditolak

Tabel 3.1. Kriteria Korelasi Point Biserial

### 3.1.4. Contoh Perhitungan Validitas dengan Point Biserial

Diketahui tabel suatu hasil jawaban siswa ditunjukkan pada tabel 3.2

:

No	Nomor Butir Instrumen												$X_i$	$X_i^2$
	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														
q														

Tabel 3.2. Contoh Hasil Jawaban Siswa

Kita coba mencari validitas soal nomor 1.

$$1) \sum X_t = 97, \quad \sum X_t^2 = 703$$

2) Nilai  $p$  = jumlah yang menjawab benar nomor 1 dibagi jumlah seluruh siswa, berarti  $p = 7 / 15 = 0,47$

$$3) q = 1 - p, \text{ berarti } q = 1 - 0,47 = 0,53$$

$$4) M_t = \frac{\sum X_t}{N} = \frac{97}{15} = 6,46$$

- 5)  $M_p$  = skor total siswa yang menjawab soal nomor 1 dengan benar dibagi jumlah siswa yang menjawab benar soal nomor 1, berarti
- $$M_p = (5+8+8+9+8+7+12) / 7 = 8,14$$

$$6) S = \sqrt{\frac{\sum X_t^2}{N} - \left(\frac{\sum X_t}{N}\right)^2}$$

$$7) S = \sqrt{\frac{703}{15} - \left(\frac{97}{15}\right)^2}$$

$$8) S = \sqrt{46,87 - 41,82} = \mathbf{2,247}$$

$$9) \text{ Jadi validitas butir soal nomor 1} = r_{pbis} = \frac{M_p - M_t}{S} \times \sqrt{\frac{p}{q}}$$

$$10) r_{pbis} = \frac{8,14 - 6,46}{2,247} \times \sqrt{\frac{0,47}{0,53}}$$

$$11) r_{pbis} = (0,748)(0,942) = \mathbf{0,704}$$

- 12) Jika dilihat dengan tabel korelasi point biserial, maka soal nomor 1 merupakan soal yang **valid**.

## 3.2. Reliabilitas Tes

### 3.2.1. Pengertian Reliabilitas

Reliabilitas berhubungan dengan masalah kepercayaan, suatu tes dapat dikatakan mempunyai kepercayaan yang tinggi jika tes itu dapat memberikan hasil yang tetap. Instrumen yang baik adalah instrumen yang dapat dengan ajeg memberikan data yang sesuai dengan kenyataan.

Reliabilitas sering diartikan dengan keterandalan atau keajegan. Artinya suatu tes memiliki keterandalan bilamana tes tersebut dipakai mengukur berulang-ulang hasilnya sama. Dengan demikian reliabilitas dapat pula diartikan dengan keajegan atau stabilitas.

Yang sering kurang tepat ditangkap pembaca adalah adanya pendapat bahwa “ajeg” atau “tetap” diartikan sebagai “sama”. Ajeg atau tetap tidak selalu harus sama, tetapi mengikuti perubahan secara ajeg. Misal, jika keadaan si A mula – mula berada lebih rendah dari si B, maka jika diadakan pengukuran ulang, si A juga berada lebih rendah dari si B. Itulah yang dikatakan ajeg atau tetap, yaitu sama antara kedudukan siswa diantara kelompok lain. Besarnya ketetapan itulah menunjukkan tingginya reliabilitas instrument.

Sehubungan dengan reabilitas ini, Scarvia B. Anderson dan kawan – kawan (1976:425) menyatakan bahwa persyaratan bagi tes, yaitu validitas dan reliabilitas ini sangat penting. Dalam hal ini validitas lebih penting dan reliabilitas ini perlu karena menyokong terbentuknya validitas. Sebuah tes mungkin reliabel tapi tidak valid. Sebaliknya, sebuah tes yang valid biasanya reliable. Sekali lagi reliabilitas adalah ketetapan suatu tes apabila diteskan kepada subjek yang sama.

### **3.2.2. Faktor yang Mempengaruhi Reliabilitas**

Ada beberapa faktor yang mempengaruhi reliabilitas,

diantaranya yaitu :

- a. Luas tidaknya sampling yang diambil

Semakin luas suatu sampling, berarti tes semakin handal.

- b. Perbedaan bakat dan kemampuan murid yang di tes

Makin variabel kemampuan peserta tes, berarti makin tinggi keandalan koefisien tes. Tes yang diberikan kepada beberapa tingkat kelas yang



berbeda lebih tinggi keandalannya daripada yang hanya diberikan kepada beberapa kelas yang sama karena tingkat kelas yang berbeda akan menghasilkan *achievement* yang lebih luas.

c. Suasana dan kondisi testing

Suasana ketika sedang berlangsung testing, seperti tenang, gaduh, banyak gangguan, pengetes yang marah-marah dapat mengganggu pengerjaan tes sehingga dengan demikian mempengaruhi pula hasil dan keandalan tes.

### 3.2.3. Teknik Pengujian Reliabilitas

Ada beberapa teknik dalam mencari reliabilitas :

a. Metode bentuk parallel (*Equivalent*)

Tes parallel adalah dua buah tes yang memiliki kesamaan tujuan, tingkat kesukaran, dan susunan, tetapi butir – butir soalnya beda. Dengan metode bentuk parallel ini, dua buah tes yang parallel, misalnya tes Matematika seri A yang akan dicari reliabilitasnya dan tes Matematika seri B diteskan kepada sekelompok siswa yang sama, kemudian hasilnya dikorelasikan. Koefisien korelasi dari kedua hasil tes inilah yang menunjukkan koefisien reliabilitas tes seri A. Jika koefisien tinggi maka tes tersebut sudah reliable dan dapat digunakan sebagai alat pengetes yang terandalkan.

Kelemahan metode ini yaitu bahwa pekerjaan dari pengetes berat karena harus menyusun dua tes seri. Lagipula harus tersedia waktu yang lama untuk mencobakan dua kali tes.

b. Metode tes ulang (*Test – Retest Method*)

Metode tes ulang dilakukan untuk menghindari penyusunan dua seri tes. Dalam melakukan metode ini pengetes hanya memiliki satu seri tes, tetapi dicobakan dua kali. Kemudian hasil dari kedua kali tes tersebut dihitung korelasinya. Pada umumnya hasil tes yang kedua lebih baik dari yang pertama. Hal ini tidak mengapa karena pengetes harus sadar akan adanya *practice effect* dan *carry over effect*. Yang penting adalah adanya kesejajaran hasil atau ketetapan hasil yang ditunjukkan oleh koefisien korelasi yang tinggi. Metode ini disebut juga korelasi diri sendiri karena mengkorelasikan hasil dari tes yang pertama.

c. Metode belah dua (*Split half method*)

Dalam menggunakan metode ini pengetes hanya menggunakan sebuah tes dan dicobakan satu kali. Pada waktu membelah dua dan mengkorelasikan dua belahan baru diketahui reliabilitas separo tes. Untuk mengetahui reliabilitas seluruh tes harus digunakan rumus Spearman-Brown sebagai berikut :

$$r_{11} = \frac{2r_{1/2,1/2}}{(1 + r_{1/2,1/2})}$$

Dimana :

$r_{1/2,1/2}$  = Korelasi antara skor – skor setiap belahan tes.

$r_{11}$  = Koefisien reliabilitas yang sudah disesuaikan.

Ada dua cara membelah butir soal ini, yaitu :

1. Membelah berdasarkan item – item genap dan ganjil yang selanjutnya disebut belahan ganjil – genap.
2. Membelah berdasarkan item item awal dan item – item akhir yaitu separo jumlah nomor – nomor awal dan separo jumlah nomor – nomor akhir yang selanjutnya disebut belahan awal – akhir.

### 3.2.4. Uji Reliabilitas

Terdapat beberapa formula untuk menghitung nilai koefisien reliabilitas yang bergantung kepada metode atau teknik pengumpulan data reliabilitas yang digunakan.

#### a) Formula KR-20

Metode KR-20 merupakan koefisien reliabilitas yang dapat menggambarkan variasi dari item – item untuk jawaban benar / salah yang diberi skor 0 atau 1. (Guilford, 1978)

Berikut merupakan rumus formula KR-20 (Surapranata, 2010:170) :

$$r_{11} = \left( \frac{n}{n-1} \right) \left( \frac{S^2 - \sum pq}{S^2} \right)$$

#### Keterangan :

$r_{11}$  : Koefisien Reliabilitas

$n$  : Banyaknya siswa

$p$  : Proporsi subjek menjawab soal dengan benar

$q$  : proporsi subjek menjawab soal dengan salah (1-p)

$S^2$  : Simpangan baku (varians)

Sedangkan rumus Simpangan Baku itu sendiri adalah :

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n}$$

Keterangan :

- $S^2$  : Simpangan baku  
 $X$  : Skor yang diperoleh siswa  
 $n$  : Banyaknya siswa

#### b) Formula KR-21

Berikut merupakan rumus formula KR-21 beserta tabel tingkat nilai reliabilitas (Arikunto, 2006:189,276).

$$r_{11} = \left( \frac{k}{k-1} \right) \left( 1 - \frac{M(k-M)}{k \times V_t} \right)$$

Keterangan :

- $r_{11}$  : Reliabilitas instrument  
 $k$  : banyaknya butir soal  
 $M$  : Skor rata – rata  
 $V_t$  : Varians total

Sedangkan untuk mendapat varians total itu sendiri adalah :

$$V_t = \sqrt{\frac{(\sum X^2) - \frac{(\sum X)^2}{n}}{n}}$$

Jika  $r_{hitung} > r_{tabel}$ , hal ini menunjukkan bahwa koefisien ada artinya hingga tidak diabaikan. Artinya instrument ini reliabel pada

taraf yang telah ditentukan yaitu 95% (Arikunto, 2006:184). Adapun nilai koefisien korelasi tingkat reliabilitas ditunjukkan pada tabel 3.3.

<b>Koefisien Korelasi (r)</b>	<b>Tafsiran</b>
$0,8 \leq r \leq 1,0$	Reliabilitas sangat tinggi
$0,6 \leq r < 0,8$	Reliabilitas tinggi
$0,4 \leq r < 0,6$	Reliabilitas sedang
$0,2 \leq r < 0,4$	Reliabilitas rendah
$r < 0,2$	Reliabilitas sangat rendah

**Tabel 3.3. Koefisien Korelasi Reliabilitas**

### **3.3. Tingkat Kesukaran**

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya, soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauan kemampuannya.

Bilangan yang menunjukkan sulit dan mudahnya suatu soal disebut dengan Indeks Kesukaran (*difficult index*). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,00. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks 0,00 menunjukkan bahwa soal terlalu sukar dan sebaliknya indeks 1,00 menunjukkan bahwa soal terlalu mudah (Arikunto, 2002:214).

Di dalam istilah evaluasi, indeks kesukaran ini diberi symbol P (p besar), singkatan dari kata ‘proporsi’.



kelompok atas maupun kelompok bawah, maka soal itu tidak baik karena tidak mempunyai daya pembeda.

Jika seluruh kelompok atas dapat menjawab soal itu dengan benar sedangkan kelompok bawah menjawab salah, maka soal tersebut memiliki nilai D paling besar yaitu 1,00. Sebaliknya jika seluruh kelompok bawah menjawab soal itu dengan benar sedangkan kelompok atas menjawab salah, maka soal itu mempunyai nilai D yaitu -1,00. Tetapi jika siswa kelompok atas dan bawah sama – sama menjawab benar atau salah, maka soal tersebut mempunyai nilai D 0,00 karena tidak mempunyai daya pembeda sama sekali.

Rumus untuk menentukan indeks diskriminasi adalah (Arikunto, 2006:295)

:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Dimana :

J : jumlah peserta tes

$J_A$  : banyaknya peserta kelompok atas

$J_B$  : banyaknya peserta kelompok bawah

$B_A$  : banyaknya peserta kelompok atas yang menjawab soal dengan benar

$B_B$  : banyaknya peserta kelompok bawah yang menjawab soal dengan benar

$P_A$  : proporsi peserta kelompok atas yang menjawab soal dengan benar

$P_B$  : proporsi kelompok bawah yang menjawab soal dengan benar

Klasifikasi daya pembeda menurut Arikunto (2006) :

D : 0,00 - 0,20 : jelek (*poor*)

D : 0,21 - 0,40 : cukup (*satisfactory*)

D : 0,41 - 0,70 : baik (*good*)

D : 0,71 - 1,00 : baik sekali (*excellent*)